

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:53:27 ; Search time 13.3 Seconds

(without alignments)
1251.566 Million cell updates/sec

Title: US-09-846-512-12

Perfect score: 2447

Sequence: 1 MGENDPPAVEAPFSEFSLFG.....TRVTSFLDWIHEQWERLKT 454

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

al number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 2447 | 100.0 | 454 | 1 | TMS3_HUMAN |
| 2 | 865.5 | 35.4 | 490 | 1 | Q91972 mus musculus |
| 3 | 864.5 | 35.3 | 492 | 1 | TMS2_HUMAN |
| 4 | 704 | 28.8 | 455 | 1 | TMS5_MOUSE |
| 5 | 696 | 28.4 | 457 | 1 | TMS5_HUMAN |
| 6 | 664.5 | 27.2 | 417 | 1 | HEPS_HUMAN |
| 7 | 661.5 | 27.0 | 1069 | 1 | ENTK_MOUSE |
| 8 | 660.5 | 27.0 | 437 | 1 | TMS4_HUMAN |
| 9 | 660 | 27.0 | 1035 | 1 | ENTK_BOVIN |
| 10 | 657 | 26.8 | 1019 | 1 | ENTK_HUMAN |
| 11 | 654 | 26.7 | 1034 | 1 | ENTK_PIG |
| 12 | 650 | 26.6 | 416 | 1 | HEPS_HUMAN |
| 13 | 645 | 26.4 | 416 | 1 | HEPS_RAT |
| 14 | 570 | 23.3 | 638 | 1 | KAL_RAT |
| 15 | 561 | 22.9 | 638 | 1 | KAL_MOUSE |
| 16 | 540 | 22.1 | 638 | 1 | KAL_HUMAN |
| 17 | 533 | 21.8 | 855 | 1 | SP14_HUMAN |
| 18 | 532 | 21.7 | 324 | 1 | TEST_MOUSE |
| 19 | 526 | 21.5 | 790 | 1 | PLMN_PIG |
| 20 | 523.5 | 21.4 | 761 | 1 | NEPR_MOUSE |
| 21 | 520.5 | 21.3 | 338 | 1 | PLMN_HORSE |
| 22 | 520 | 21.3 | 1042 | 1 | CORI_HUMAN |
| 23 | 512.5 | 20.9 | 855 | 1 | ST14_MOUSE |
| 24 | 508.5 | 20.8 | 314 | 1 | TEST_HUMAN |
| 25 | 507.5 | 20.7 | 812 | 1 | PLMN_MOUSE |
| 26 | 507 | 20.7 | 437 | 1 | ACRO_RAT |
| 27 | 506 | 20.7 | 1113 | 1 | CORI_MOUSE |
| 28 | 504.5 | 20.6 | 810 | 1 | PLMN_HUMAN |
| 29 | 503.5 | 20.6 | 273 | 1 | MCT7_MOUSE |
| 30 | 503.5 | 20.6 | 436 | 1 | ACRO_MOUSE |
| 31 | 503 | 20.6 | 343 | 1 | PLMN_SHEEP |
| 32 | 501.5 | 20.5 | 810 | 1 | PLMN_SHEEP |
| 33 | 501.5 | 20.5 | 875 | 1 | NEPR_HUMAN |

| | | | | | | |
|----|-------|------|-----|---|------------|--------------------|
| 34 | 500.5 | 20.5 | 275 | 1 | TRYT_CANFA | P15944 canis faml |
| 35 | 500.5 | 20.5 | 421 | 1 | ACRO_HUMAN | P10323 homo sapien |
| 36 | 499.5 | 20.4 | 342 | 1 | PS88_RAT | P06887 rattus norv |
| 37 | 498.5 | 20.4 | 431 | 1 | ACRO_RABIT | P48038 oryctolagus |
| 38 | 498.5 | 20.4 | 812 | 1 | PLMN_BOVIN | P06868 bos taurus |
| 39 | 497.5 | 20.3 | 810 | 1 | PLMN_MACMU | P12545 macaca mla |
| 40 | 495.5 | 20.2 | 270 | 1 | TRYT_MERIN | P50342 meriones un |
| 41 | 495.5 | 20.2 | 333 | 1 | PLMN_CANFA | P80009 canis faml |
| 42 | 495.5 | 20.2 | 342 | 1 | PS88_MOUSE | P068d1 mus musculu |
| 43 | 494.5 | 20.2 | 625 | 1 | FA11_HUMAN | P21845 mus musculu |
| 44 | 490.5 | 20.0 | 276 | 1 | MCT6_MOUSE | P21845 mus musculu |
| 45 | 488.5 | 20.0 | 274 | 1 | TRYM_RAT | P50343 rattus norv |

ALIGNMENTS

RESULT 1
TMS3_HUMAN STANDARD: PRT: 454 AA.
ID P57727;
AC 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSMEMBRANE PROTEASE, SERINE 3 (EC 3.4.21.-) (SERINE PROTEASE
TAG-12) (TUMOR ASSOCIATED DIFFERENTIALLY-EXPRESSED GENE-12 PROTEIN).
GN TMPS3 OR TADG12 OR ECHOS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A AND TRUNCATED).
RC TISSUE=Ovarian carcinoma;
RX MEDLINE=20521358; PubMed=11068177;
RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
Wang Y., Paroley T.H., O'Brien T.J.;
RT "Ovarian tumor cells express a novel multi-domain cell surface serine
protease.";
RL Biochim. Biophys. Acta 1502:337-350(2000).
[2]
RP SEQUENCE FROM N.A. (ISOFORMS A; B/C AND D), AND VARIANT ILE-53.
RX MEDLINE=20578749; PubMed=11137999;
RA Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chrast R.,
Giupponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
Younus F., Mohd S.O., Radhakrishna U., Papasavvas M.P., Gehrig C.,
Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B.,
Antonarakis S.E.;
RT "Insertion of beta-satellite repeats identifies a transmembrane
protease causing both congenital and childhood onset autosomal
recessive deafness.";
RL Nat. Genet. 27:59-63(2001).
CC -1- FUNCTION: PROBABLE PROTEASE.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/C, D AND
TRUNCATED/TADG-12V; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. THE TRUNCATED
ISOFORM IS FOUND AT INCREASED LEVELS IN SOME CARCINOMAS.
CC -1- DISEASE: DEFECTS IN TMPS3 ARE A CAUSE OF TWO FORMS OF AUTOSOMAL
NEUROSENSORY CHILDHOOD-ONSET FORMS OF DEAFNESS, DFNB8 AND DFNB10.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
or send an email to license@isb.sib.ch).

DR EMBL: AF201380; AAC37012.1; -
 DR EMBL: AB038157; BAB20077.1; -
 DR EMBL: AB038158; BAB20078.1; -
 DR EMBL: AB038159; BAB20079.1; -
 DR EMBL: AB038160; BAB20080.1; -
 DR MIM: 605511; -
 DR MIM: 601072; -
 DR MIM: 605316; -
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LDL_recept_A.
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00057; ldl_recept_a; 1.
 DR Pfam: PF00530; SRCR; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR SMART: SM00192; ldl; 1.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00202; TRYP_SPE; 1.
 DR PROSITE: PS01209; LDLRA_1; 1.
 DR PROSITE: PS00068; LDLRA_2; 1.
 DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS00287; SRCR_2; 1.
 DR PROSITE: PS02040; TRYPIN_DOM; 1.
 DR PROSITE: PS00134; TRYPIN_HIS; 1.
 DR PROSITE: PS00135; TRYPIN_SER; 1.
 DR Hydrolase: Serine protease; Transmembrane; Signal-anchor; Deafness;
 KW Alternative splicing; Polymorphism.
 KW CYTOPLASMIC (POTENTIAL).
 KW SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 KW EXTRACELLULAR (POTENTIAL).
 KW LDL-RECEPTOR CLASS A. SRCR.
 FT DOMAIN 70 454
 FT DOMAIN 72 108
 FT DOMAIN 109 205
 FT DOMAIN 217 454
 FT ACT_SITE 257 257
 FT ACT_SITE 304 304
 FT ACT_SITE 401 401
 FT SITE 216 217
 FT DISULFID 73 85
 FT DISULFID 79 98
 FT DISULFID 92 107
 FT DISULFID 129 194
 FT DISULFID 142 204
 FT DISULFID 207 324
 FT DISULFID 242 258
 FT DISULFID 338 407
 FT DISULFID 370 386
 FT DISULFID 397 425
 FT CARBOHYD 221 221
 FT VARSPIC 1 127
 FT VARSPIC 318 454
 FT VARSPIC 261 293
 FT VARSPIC 294 454
 FT VARIANT 53 53
 FT CONFLICT 46 54
 FT CONFLICT 90 90
 FT CONFLICT 350 350
 FT CONFLICT 369 395
 FT CONFLICT 427 437
 FT SEQUENCE 454 AA; 49404 MM; 57EC3678F7D6AFC CRC64;
 Query Match 100.0%; Score 2447; DB 1; Length 454;
 Best Local Similarity 100.0%; Pred. No. 6,9e-201;
 Matches 454; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGENDPRAVEAFPSRSLFGDLDTKISVPADADVAQAIIISLPLKFFPIIVIGIIALI 60
 Db 1 MGENDPRAVEAFPSRSLFGDLDTKISVPADADVAQAIIISLPLKFFPIIVIGIIALI 60
 QY 61 LALAIGLGHPCSCGKRCRSRSEKIELIARCDGVSDCKDGEDERCVRVGQNNVLOVF 120
 Db 61 LALAIGLGHPCSCGKRCRSRSEKIELIARCDGVSDCKDGEDERCVRVGQNNVLOVF 120
 QY 121 TAASAKTKMSDDMKHYANVACAQGFPSYSSDMLRVSSLEGQPREFEVSDIHLPLPDK 180
 Db 121 TAASAKTKMSDDMKHYANVACAQGFPSYSSDMLRVSSLEGQPREFEVSDIHLPLPDK 180
 QY 181 VTALHSHVYVRGCGASGHVYVLTCTACGHRGYSRIRYGGNNLSLQPMQASLOFGYH 240
 Db 181 VTALHSHVYVRGCGASGHVYVLTCTACGHRGYSRIRYGGNNLSLQPMQASLOFGYH 240
 QY 241 LCGGSVTPPLWITTAHCVYDYLTPKSWTIQVGLVSLDNPAPSHLVEKIVYHSKPKPR 300
 Db 241 LCGGSVTPPLWITTAHCVYDYLTPKSWTIQVGLVSLDNPAPSHLVEKIVYHSKPKPR 300
 QY 301 LGNDIALMKLAGPLTFENEMIQVCLPNSSENPDDKVCWTSQMGATEDGAGDASPVLANH 360
 Db 301 LGNDIALMKLAGPLTFENEMIQVCLPNSSENPDDKVCWTSQMGATEDGAGDASPVLANH 360
 QY 361 AVPLISNKCINHRDVGIGIISPSMLCAGYLTGVDSCGSGGLVCOERRLMKLVGATS 420
 Db 361 AVPLISNKCINHRDVGIGIISPSMLCAGYLTGVDSCGSGGLVCOERRLMKLVGATS 420
 QY 421 FGIGCAEVNKPQVTVRYVSLDNIHOMERDCKT 454
 Db 421 FGIGCAEVNKPQVTVRYVSLDNIHOMERDCKT 454
 QY 421 FGIGCAEVNKPQVTVRYVSLDNIHOMERDCKT 454
 Db 421 FGIGCAEVNKPQVTVRYVSLDNIHOMERDCKT 454
 RESULT 2
 TMS2_MOUSE STANDARD: PRT: 490 AA.
 ID TMS2_MOUSE
 AC Q9J0B8; Q9J0K4; Q9J082;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSMEMBRANE PROTEIN, SERINE 2 (EC 3.4.21.-) (EPITHELIN) (PLASMIC
 DE TRANSMEMBRANE PROTEIN X).
 GN TMPSR2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_Taxid=10090;
 RN RN
 RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RN RC STRAIN-BALB/C;
 RN RX MEDLINE-21104370; PubMed-11169526;
 RN RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikio P.T.;
 RN "Expression of transmembrane serine protease TMPSR2 in mouse and
 RN human tissues";
 RN J. Pathol. 193:134-140(2001).
 RN [2]
 RN SEQUENCE FROM N.A.
 RN RA Han J., Kim S.;
 RN "Putative transmembrane protease X";
 RN Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RN RC STRAIN-BALB/C;
 RN RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
 RN "A novel mosaic serine protease, epithelisin";
 RN Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
 RN RL -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN PROSTATE AND KIDNEY.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.

PR 29-APR-1998; 9805-008349

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:52:32 ; Search time 15.91 seconds
(without alignments)
642.143 Million cell updates/sec

Title: US-09-846-512-12
Perfect score: 2447
Sequence: 1 MGENDPPEAVEAFSFRSLFG.....TRVTSFLDWIHEQMERDLKT 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents-AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-------------------|--------------------|
| 1 | 2375 | 97.1 | 454 | US-09-518-046-2 | Sequence 2, Appl1 |
| 2 | 2195.5 | 89.7 | 455 | US-09-261-416-2 | Sequence 2, Appl1 |
| 3 | 1354 | 55.3 | 294 | US-09-518-046-4 | Sequence 4, Appl1 |
| 4 | 875.5 | 35.8 | 492 | US-09-342-749-2 | Sequence 2, Appl1 |
| 5 | 851.5 | 34.8 | 157 | US-09-518-046-23 | Sequence 23, Appl1 |
| 6 | 686.5 | 28.1 | 283 | US-08-807-151-1 | Sequence 1, Appl1 |
| 7 | 660.5 | 27.0 | 435 | US-09-008-271A-6 | Sequence 6, Appl1 |
| 8 | 660 | 27.0 | 798 | US-08-200-900A-2 | Sequence 2, Appl1 |
| 9 | 660 | 27.0 | 416 | US-09-000-846-2 | Sequence 2, Appl1 |
| 10 | 655.5 | 26.8 | 416 | US-09-000-846-2 | Sequence 2, Appl1 |
| 11 | 576 | 23.5 | 255 | US-09-027-337-3 | Sequence 3, Appl1 |
| 12 | 571 | 23.3 | 255 | US-08-944-483-67 | Sequence 67, Appl1 |
| 13 | 570 | 23.3 | 638 | US-08-681-151-3 | Sequence 3, Appl1 |
| 14 | 564.5 | 23.1 | 418 | US-08-508-448C-25 | Sequence 25, Appl1 |
| 15 | 533 | 21.8 | 232 | US-08-508-448C-19 | Sequence 19, Appl1 |
| 16 | 533 | 21.6 | 855 | US-09-027-337-2 | Sequence 2, Appl1 |
| 17 | 529.5 | 21.5 | 248 | US-08-944-483-63 | Sequence 63, Appl1 |
| 18 | 527 | 21.5 | 235 | US-08-807-151-3 | Sequence 3, Appl1 |
| 19 | 523 | 21.4 | 98 | US-09-518-046-17 | Sequence 17, Appl1 |
| 20 | 518 | 21.2 | 235 | US-08-944-483-65 | Sequence 65, Appl1 |
| 21 | 509.5 | 20.8 | 546 | 5200340-6 | Patent No. 5200340 |
| 22 | 508.5 | 20.8 | 314 | US-09-008-271A-3 | Sequence 3, Appl1 |
| 23 | 507.5 | 20.7 | 812 | US-08-248-629A-1 | Sequence 1, Appl1 |
| 24 | 507.5 | 20.7 | 812 | US-08-451-932-1 | Sequence 1, Appl1 |
| 25 | 507.5 | 20.7 | 812 | US-08-452-260-1 | Sequence 1, Appl1 |
| 26 | 507.5 | 20.7 | 812 | US-08-326-785-1 | Sequence 1, Appl1 |
| 27 | 507.5 | 20.7 | 812 | US-08-612-788-1 | Sequence 1, Appl1 |

| | | | | | |
|----|-------|------|-----|-------------------|--------------------|
| 28 | 507.5 | 20.7 | 812 | US-08-605-598B-1 | Sequence 1, Appl1 |
| 29 | 507.5 | 20.7 | 812 | US-08-429-743-1 | Sequence 1, Appl1 |
| 30 | 507.5 | 20.7 | 812 | US-08-866-743-1 | Sequence 1, Appl1 |
| 31 | 507.5 | 20.7 | 812 | US-09-066-028-1 | Sequence 1, Appl1 |
| 32 | 507.5 | 20.7 | 812 | PCT-US95-05107-1 | Sequence 1, Appl1 |
| 33 | 507 | 20.7 | 814 | US-08-750-711-1 | Sequence 1, Appl1 |
| 34 | 505.5 | 20.7 | 791 | US-08-643-219-1 | Sequence 1, Appl1 |
| 35 | 505.5 | 20.7 | 791 | US-08-851-350-1 | Sequence 1, Appl1 |
| 36 | 504.5 | 20.6 | 790 | US-08-469-486-54 | Sequence 54, Appl1 |
| 37 | 504.5 | 20.6 | 790 | US-08-469-658-54 | Sequence 54, Appl1 |
| 38 | 504.5 | 20.6 | 791 | US-09-131-995-1 | Sequence 1, Appl1 |
| 39 | 504.5 | 20.6 | 791 | US-08-832-087B-1 | Sequence 1, Appl1 |
| 40 | 504.5 | 20.6 | 791 | US-09-132-154-1 | Sequence 1, Appl1 |
| 41 | 504.5 | 20.6 | 810 | US-07-854-603-2 | Sequence 2, Appl1 |
| 42 | 504.5 | 20.6 | 810 | US-08-147-000B-29 | Sequence 29, Appl1 |
| 43 | 504.5 | 20.6 | 810 | US-09-086-514-1 | Sequence 1, Appl1 |
| 44 | 504.5 | 20.6 | 810 | 5200340-8 | Patent No. 5200340 |
| 45 | 503.5 | 20.6 | 273 | US-08-978-404B-3 | Sequence 3, Appl1 |

ALIGNMENTS

RESULT 1
US-09-518-046-2
; Sequence 2, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; EARLIER FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRP
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TAD6-12
; OTHER INFORMATION: protein
US-09-518-046-2

| | | | | |
|-----------------------|--|--------------------|------------|------------|
| Query Match | 97.1% | Score 2375; | DB 4; | Length 454 |
| Best Local Similarity | 97.6% | Pred No. 1.5e-224; | | |
| Matches 444; | Conservative | 2; | Mismatches | Indels 2; |
| | | | Gaps | 2; |
| QY 1 | MGENDPPEAVEAFSFRSLFGDLDTKISPAADAVAAQILSLPLKPF-PIIVIGIALL | 59 | | |
| DB 1 | MGENDPPEAVEAFSFRSLFGDLDTKISPAADAVAAQILSLPLPREVSQSSSIGIALL | 60 | | |
| QY 60 | ILALAIIGLGIHDDCSKRYKCRSSFKCIELIARCDNSDKDQEDERYCYRVGONAAV | 119 | | |
| DB 61 | ILALAIIGLGIHDDCSKRYKCRSSFKCIELITRCDGSDCKDDEDEKCYRVGONAAV | 120 | | |
| QY 120 | PTAASWKTCSDMKGHNVAACAOIGFSPYSSDNLRVSSLEGOPREFEVSIDHLLPPD | 179 | | |
| DB 121 | PTAASWKTCSDMKGHNVAACAOIGFSPYSSDNLRVSSLEGOPREFEVSIDHLLPPD | 180 | | |
| QY 180 | KYTAALHSHVYVBEGCASGHVVTLOCTACGHRGRGYSRIIVGNMSSLQMPQASLOFGY | 239 | | |
| DB 181 | KYTAALHSHVYVBEGCASGHVVTLOCTACGHRGRGYSRIIVGNMSSLQMPQASLOFGY | 240 | | |
| QY 240 | HICGGSVITPLKLTITAAHGVYDLYLPKSWTIOGVLSILNDNAPSHLYEKITYHSHKYPK | 299 | | |
| DB 241 | HICGGSVITPLKLTITAAHGVYDLYLPKSWTIOGVLSILNDNAPSHLYEKITYHSHKYPK | 300 | | |
| QY 300 | RLGNDIALMKLAGPLTFNEMIOVCLPNSSENFPGKVCWTSGMGATDGDAGDASPLVNH | 359 | | |

```

Db 301 RLGNDAIMKLAGPLTFNEMIQPCLPNSSENEFPDGKVCWTSGMATEDG-CDASPYLNH 359
QY 360 AAVPLISNKCINHRDVGGLIISPSMCAAGLTGCVDSOCODSGPLVCORRLMKLVAT 419
Db 360 AAVPLISNKCINHRDVGGLIISPSMCAAGLTGCVDSOCODSGPLVCORRLMKLVAT 419
QY 420 SFGICAEVKNKPGVYTRVTSFLDMIHQEMERDLKT 454
Db 420 SFGICAEVKNKPGVYTRVTSFLDMIHQEMERDLKT 454

```

```

RESULT 2
US-09-261-416-2
; Sequence 2, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-12 encoded by nucleotides
; Patent No. 6291663
US-09-261-416-2

```

```

Query Match 89.7%; Score 2195.5; DB 4; Length 455;
Best Local Similarity 91.5%; Pred. No. 6e-207;
Matches 421; Conservative 5; Mismatches 23; Indels 11; Gaps 4;

```

```

QY 1 MGENDPRAVAPRPSFRLFLDLTKISPAVDADAVAAQILSLPLKFF-PIIVIGIIL 59
Db 1 MGENDPRAVAPRPSFRLFLDLTKISPAVDADAVAAQILSLPLFEVFSQSSSLGIIL 60
QY 60 ILALAIGLIGHPCSGKYRCSRSEKIELTARCDGVSDCKDGEDEYRCVAVGONAVLOY 119
Db 61 ILALAIGLIGHPCSGKYRCSRSEKIELTARCDGVSDCKDGEDEYRCVAVGONAVLOY 120
QY 120 FTAASWKTMCSDMKGHYANVACAQLGFPYSVSDNLRVSSLGQFREFEVSIDHLLPDD 179
Db 121 FTAASWKTMCSDMKGHYANVACAQLGFPYSVSDNLRVSSLGQFREFEVSIDHLLPDD 180
QY 180 KYTALHHSYVRREGASGHVYTLQCTACGHRGYSRIVGNNLSLSQMPWQASLOFOGY 239
Db 181 KYTALHHSYVRREGASGHVYTLQCTACGHRGYSRIVGNNLSLSQMPWQASLOFOGY 240
QY 240 HLCGGSVITPLMTITTAHCYVDLYLPKSWTIQYGLVSLDNPAPSHLYEIKIVHSKKPK 299
Db 241 HLCGGSVITPLMTITTAHCYVDLYLPKSWTIQYGLVSLDNPAPSHLYEIKIVHSKKPK 300
QY 300 RLGNDAIMKLAGPLTFNEMIQPCLPNSSENEFPDGKVCWTSGMATEDGAGDASPYLNH 359
Db 301 RLGNDAIMKLAGPLTFNEMIQPCLPNSSENEFPDGKVCWTSGMATEDG-CDASPYLNH 359
QY 360 AAVPLISNKCINHRDVGGLIISPSMCAAGLTGCVDSOCODSGPLVCORRLMKLVAT 419
Db 360 AAVPLISNKCINHRDVGGLIISPSMCAAGLTGCVDSOCODSGPLVCORRLMKLVAT 419
QY 415 LVGATSEFGICAEVKNKPGVYTRVTSFLDMIHQEMERDLKT 454
Db 416 LVGATSEFGICAEVKNKPGVYTRVTSFLDMIHQEMERDLKT 455

```

```

RESULT 3
US-09-518-046-4
; Sequence 4, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 4
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
US-09-518-046-4

```

```

Query Match 55.3%; Score 1354; DB 4; Length 294;
Best Local Similarity 88.1%; Pred. No. 9.7e-125;
Matches 258; Conservative 7; Mismatches 20; Indels 8; Gaps 2;

```

```

QY 1 MGENDPRAVAPRPSFRLFLDLTKISPAVDADAVAAQILSLPLKFF-PIIVIGIIL 59
Db 1 MGENDPRAVAPRPSFRLFLDLTKISPAVDADAVAAQILSLPLFEVFSQSSSLGIIL 60
QY 60 ILALAIGLIGHPCSGKYRCSRSEKIELTARCDGVSDCKDGEDEYRCVAVGONAVLOY 119
Db 61 ILALAIGLIGHPCSGKYRCSRSEKIELTARCDGVSDCKDGEDEYRCVAVGONAVLOY 120
QY 120 FTAASWKTMCSDMKGHYANVACAQLGFPYSVSDNLRVSSLGQFREFEVSIDHLLPDD 179
Db 121 FTAASWKTMCSDMKGHYANVACAQLGFPYSVSDNLRVSSLGQFREFEVSIDHLLPDD 180
QY 180 KYTALHHSYVRREGASGHVYTLQCTACGHRGYSRIVGNNLSLSQMPWQASLOFOGY 239
Db 181 KYTALHHSYVRREGASGHVYTLQCTACGHRGYSRIVGNNLSLSQMPWQASLOFOGY 240
QY 240 HLCGGSVITPLMTITTAHCYVDLYLPKSWTIQYGLVSLDNPAPSH 285
Db 241 HLCGGSVITPLMTITTAHCYVDLYLPKSWTIQYGLVSLDNPAPSH 293

```

```

RESULT 4
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 Is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

```


DB 250 TSMGSCAKARPGYGNVWFTDWIYROMRAD 282

RESULT 7

US-09-008-271A-6

Sequence 6, Application US/09008271A

Patent No. 6203979

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Hillman, Jennifer L.

Yue, Henry

Guegler, Karl J.

Corley, Neil C.

Tang, Tom Y.

Shah, Puryi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastBio for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/008,271A

FILING DATE: 16-Jan-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: <Unknown>

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mohan-Peterson, Sheila

REGISTRATION NUMBER: 41,201

REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 435 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLNOT13

CLONE: 1337018

SEQUENCE DESCRIPTION: SEQ ID NO: 6 :

US-09-008-271A-6

Query Match 27.0%; Score 660.5; DB 4; Length 435;

Best Local Similarity 35.4%; Pred. No. 1.6e-56;

Matches 157; Conservative 73; Mismatches 150; Indels 63; Gaps 17;

QY 44 LPIKFPPIVIGIILALALA-----IGGIHFDCSGKYRCRSSFKC--IELIAR-- 91

DB 22 IPMEFRKVGIPILIALISILIVVLKILVD--KY----YFLCGGPHLHPIRKOL 74

QY 92 CDSVSGCKOGEDEYRC-----VRGGGNAYLVQVTTAA--SKTKCSDDMKGHYA 138

DB 75 CDSGLDCCPLGEDEHCKVSPGPAVAVRLSKRSTLVQVDSATGMWFSACFTNFTAL 134

QY 139 NVACAOLGFPYSVSDNLKRVSLGQFREFEVSIDHLLPDKYATLHHSYVREG--CA 195

DB 135 ETACRQMGYS-----KPIFRAYEIGRPDDLDVVELTENSGELRMNNSGPCL 182

QY 196 GHVVTLOCTACGHRGYSRIYVGNNSLSIQMPQASLOFGYHLCGGSVITPLMTITA 255

DB 183 SGSLVSLHCLACGESI-KTPRVVGCSEASVDPMQVSIQYDKOHGCGSILDPHVLTA 241

QY 256 AHCV---YDLXLPKSMTIOGVGLSLDNPAPSHLVKIV---VHSYKPKRLGNDIALMK 309

DB 242 AHCFRKHTDYF---NMKVRAGSDKL--GSFPLAVAKIITIEFNPV--PK--DNDIALMK 293

QY 310 LAGPLTFENEMIOVCLPNSSENEPDKVCWTSCGATGADGASPVLANAVALPISNKI 369

DB 294 LQPLTFESCTVRPCLPFDEELTPATPLIIGMGTQKMGKMSDILLQASVOYIDSTR 353

QY 370 CNHRDVGIGIISPMACAGLTGCVSCQSDSGPLVQCEERLMKLVGATSEIGCAEVN 429

DB 354 CNADDAVYQGEVTEKMKACIGPEGGVDTCCDSGGLMYOSDQ--WHVYGVYSWVGCGGPS 412

QY 430 KPGVYTRVTSFLDMIHQWERDL 452

DB 413 TPGVYTKVSAYLWMTYVWKAEL 435

RESULT 8

US-08-200-900A-2

Sequence 2, Application US/08200900A

Patent No. 5665566

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/200,900A

FILING DATE: 23-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Meinhert, Maureen C.

REGISTRATION NUMBER: 31,544

REFERENCE/DOCKET NUMBER: GI 5201-FWC

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170 X8574

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 798 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-200-900A-2

Query Match 27.0%; Score 660; DB 1; Length 798;

Best Local Similarity 36.9%; Pred. No. 4.3e-56;

Matches 146; Conservative 70; Mismatches 146; Indels 34; Gaps 15;

QY 66 GLGIHFDC-SGKYRCSSFKCIELIARCDGVSCCKGGEDEYKRVY---GGONAVLVQV 120

DB 415 GLGIPECKEDNFQCKDG-ECIPLVNLCDGFPCKCKGSDSEAHCVRLFNGTTDSSGLVQFR 473

QY 121 TAASWTMCSDDMKGHYANVACAOLGFPYSVSDNLKRVSLGQFREFEVSIDHLLPDK 180

DB 474 IOSIMHVACENMTTQISDVCOLLGLGT--GNSSVPTSTGG--GPVYNLN----- 521

QY 181 VTALHHSYV--REGCASGHVVTLOCT--ACGHR---RGYSRIYVGNNSLSIQMPQAS 233

Db 270 EYIQVCLPAAGALVDGKVCYVIGWNTQ-FYGOAMVLQEARVPLISNEVCNSPDEY 328
QY 378 GLSPSMLCAGYLTGVDSCGDSGGLVCOE-----RLMKLVGATSFSGICAEVKNPGV 433
Db 329 NQKPKMFCAGYPEGIDACQDSGSGFVCEDSISGTSRMRKLCGYISWGTGICALARKPGV 388
QY 434 YTRVTSFLDWIHEOME 449
Db 389 YTKVTFDEREWIFKAIRK 404

RESULT 11

US-09-027-337-3
; Sequence 3, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; FILE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 3
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of hepsin (Heps)
; OTHER INFORMATION: homologous to similar domain in TADG-15
US-09-027-337-3

Query Match 23.5%; Score 576; DB 2; Length 256;
Best Local Similarity 43.1%; Pred. No. 1.4e-48;

Matches 114; Conservative 37; Mismatches 74; Indels 28; Gaps 7;

QY 216 RIYGNMSSLQMPQWASLOFOGHLGGSVITPLMTITTAHCYVDLYLPK-----SWT 269
Db 1 RIYGNMSSLQMPQWASLOFOGHLGGSVITPLMTITTAHCYVDLYLPK-----SWT 269
QY 270 IQVGLSLDNPAPSHL---VEKIVYHSKYKPKR-----LGNDIALMKLAGPLTFNEMT 320
Db 56 VFAGAVA---QASPHGLQIGVQAVVYHGVLPRDPNSENSNDIALVHLSPLPLEYI 112
QY 321 QPCLPENSENFPDGKVCWTSWGATGAGDASPVLNHAAPLISKICNHRVYGGIT 380
Db 113 QPCLPENSENFPDGKVCWTSWGATGAGDASPVLNHAAPLISKICNHRVYGGIT 380
QY 381 SPMLCAGYLTGVDSCGDSGGLVCOE-----RLMKLVGATSFSGICAEVKNPGVYTR 436
Db 172 KPMFCAGYPEGIDACQDSGSGFVCEDSISGTSRMRKLCGYISWGTGICALARKPGVYTR 231
QY 437 VTSFLDWIHEOME 449
Db 232 VSDFEREWIFKAIRK 244

RESULT 12

US-08-944-483-67
; Sequence 67, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLAS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STRODE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF INVENTION: OF THE PROSTATE
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESS: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

APPLICATION NUMBER: US/08/944,483

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:

NAME: Becker, Cheryl L.

REGISTRATION NUMBER: 35,441

REFERENCE/DOCKET NUMBER: 6183.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 847/935-1729

TELEFAX: 847/938-2623

TELEX:

INFORMATION FOR SEQ ID NO: 67:

SEQUENCE CHARACTERISTICS:

LENGTH: 255 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6232456e

US-08-944-483-67

Query Match 23.3%; Score 571; DB 4; Length 255;

Best Local Similarity 44.8%; Pred. No. 4.4e-48;

Matches 113; Conservative 37; Mismatches 74; Indels 28; Gaps 7;

QY 217 IYGNMSSLQMPQWASLOFOGHLGGSVITPLMTITTAHCYVDLYLPK-----SWT 270
Db 1 IYGNMSSLQMPQWASLOFOGHLGGSVITPLMTITTAHCYVDLYLPK-----SWT 270
QY 271 IQVGLSLDNPAPSHL---VEKIVYHSKYKPKR-----LGNDIALMKLAGPLTFNEMT 321
Db 56 VFAGAVA---QASPHGLQIGVQAVVYHGVLPRDPNSENSNDIALVHLSPLPLEYI 112
QY 322 QPCLPENSENFPDGKVCWTSWGATGAGDASPVLNHAAPLISKICNHRVYGGIT 381
Db 113 QPCLPENSENFPDGKVCWTSWGATGAGDASPVLNHAAPLISKICNHRVYGGIT 381
QY 382 SPMLCAGYLTGVDSCGDSGGLVCOE-----RLMKLVGATSFSGICAEVKNPGVYTR 437
Db 172 KPMFCAGYPEGIDACQDSGSGFVCEDSISGTSRMRKLCGYISWGTGICALARKPGVYTR 231
QY 438 TSFLDWIHEOME 449
Db 232 SDFEREWIFKAIRK 243

RESULT 13

US-08-681-151-3
; Sequence 3, Application US/08681151
; Patent No. 5869637
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Braxton, Scott Michael

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:53:27 ; Search time 13.3 Seconds
(without alignments)
1251.566 Million cell updates/sec

Title: US-09-846-512-12

Perfect score: 2447
Sequence: 1 MGENDPPAVEAPFSFSLFG.....TRVTSFLDWTHEQMERDLKT 454

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 2447 | 100.0 | 454 | 1 | TMS3_HUMAN |
| 2 | 865.5 | 35.4 | 490 | 1 | TMS2_MOUSE |
| 3 | 864.5 | 35.3 | 492 | 1 | TMS2_HUMAN |
| 4 | 704 | 28.8 | 455 | 1 | TMS2_MOUSE |
| 5 | 666 | 28.4 | 457 | 1 | TMS5_HUMAN |
| 6 | 664.5 | 27.2 | 417 | 1 | HEPS_HUMAN |
| 7 | 661.5 | 27.0 | 1069 | 1 | ENTK_MOUSE |
| 8 | 660.5 | 27.0 | 437 | 1 | TMS4_HUMAN |
| 9 | 660 | 27.0 | 1035 | 1 | ENTK_BOVIN |
| 10 | 657 | 26.8 | 1019 | 1 | ENTK_HUMAN |
| 11 | 654 | 26.7 | 1034 | 1 | ENTK_PIG |
| 12 | 650 | 26.6 | 416 | 1 | HEPS_MOUSE |
| 13 | 645 | 26.4 | 416 | 1 | HEPS_RAT |
| 14 | 570 | 23.3 | 638 | 1 | KAL_RAT |
| 15 | 561 | 22.9 | 638 | 1 | KAL_MOUSE |
| 16 | 540 | 22.1 | 638 | 1 | KAL_HUMAN |
| 17 | 533 | 21.8 | 855 | 1 | ST14_HUMAN |
| 18 | 532 | 21.7 | 324 | 1 | TEST_MOUSE |
| 19 | 526 | 21.5 | 790 | 1 | PLMN_PIG |
| 20 | 523.5 | 21.4 | 761 | 1 | NETR_MOUSE |
| 21 | 520.5 | 21.3 | 338 | 1 | PLMN_HORSE |
| 22 | 520 | 21.3 | 1042 | 1 | COR1_HUMAN |
| 23 | 512.5 | 20.9 | 855 | 1 | ST14_MOUSE |
| 24 | 508.5 | 20.8 | 314 | 1 | TEST_HUMAN |
| 25 | 507.5 | 20.7 | 812 | 1 | PLMN_MOUSE |
| 26 | 507 | 20.7 | 437 | 1 | ACRO_RAT |
| 27 | 506 | 20.7 | 1113 | 1 | COR1_MOUSE |
| 28 | 504.5 | 20.6 | 810 | 1 | PLMN_HUMAN |
| 29 | 503.5 | 20.6 | 273 | 1 | ACR7_MOUSE |
| 30 | 503.5 | 20.6 | 436 | 1 | ACR7_MOUSE |
| 31 | 503 | 20.6 | 343 | 1 | PLMN_SHEEP |
| 32 | 501.5 | 20.5 | 810 | 1 | PLMN_FRIED |
| 33 | 501.5 | 20.5 | 875 | 1 | NETR_HUMAN |

| | | | | | | |
|----|-------|------|-----|---|------------|--------------------|
| 34 | 500.5 | 20.5 | 275 | 1 | TRYT_CANFA | P15944 canis fam1 |
| 35 | 500.5 | 20.5 | 421 | 1 | ACRO_HUMAN | P10323 homo sapien |
| 36 | 499.5 | 20.4 | 342 | 1 | PSS8_RAT | O96807 rattus norv |
| 37 | 498.5 | 20.4 | 431 | 1 | ACRO_RABIT | P48038 oryctolagus |
| 38 | 498.5 | 20.4 | 812 | 1 | PLMN_BOVIN | P06868 bos taurus |
| 39 | 497.5 | 20.3 | 810 | 1 | PLMN_MACRO | P12345 macaca mula |
| 40 | 495.5 | 20.2 | 270 | 1 | TRYT_MERUN | P50342 meriones un |
| 41 | 495.5 | 20.2 | 333 | 1 | PLMN_CANFA | P80009 canis fam1 |
| 42 | 495.5 | 20.2 | 342 | 1 | PSS8_MOUSE | O96801 mus musculu |
| 43 | 494.5 | 20.2 | 625 | 1 | FAIL_HUMAN | P03951 homo sapien |
| 44 | 490.5 | 20.0 | 276 | 1 | MCT6_MOUSE | P21845 mus musculu |
| 45 | 488.5 | 20.0 | 274 | 1 | TRYM_RAT | P50343 rattus norv |

ALIGNMENTS

| RESULT | ID | TMS3_HUMAN | STANDARD | PRT | 454 AA. |
|--------|--|-----------------------------------|----------|-----|---------|
| AC | P57727 | | | | |
| DT | 20-AUG-2001 | (Rel. 40, Created) | | | |
| DT | 20-AUG-2001 | (Rel. 40, Last sequence update) | | | |
| DT | 20-AUG-2001 | (Rel. 40, Last annotation update) | | | |
| DE | TRANSMEMBRANE PROTEASE, SERINE 3 (EC 3.4.21.-) (SERINE PROTEASE | | | | |
| DE | TADG-12) (TUMOR ASSOCIATED DIFFERENTIALLY-EXPRESSED GENE-12 PROTEIN). | | | | |
| GN | TPRSS3 OR TADG12 OR ECHOS1. | | | | |
| OS | Homo sapiens (Human). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. | | | | |
| OX | NCBI_TaxID=9606; | | | | |
| ON | [1] | | | | |
| RP | SEQUENCE FROM N.A. (ISOFORMS A AND TRUNCATED). | | | | |
| RC | TISSUE=ovarian carcinoma. | | | | |
| RA | MEDLINE-20521358; PubMed-11068177; | | | | |
| RA | Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N., | | | | |
| RA | Wang Y., Parmley T.H., O'Brien T.J.; | | | | |
| RT | "Ovarian tumor cells express a novel multi-domain cell surface serine | | | | |
| RT | protease". | | | | |
| RL | Biochim. Biophys. Acta 1502:337-350(2000). | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N.A. (ISOFORMS A; B/C AND D), AND VARIANT ILE-53. | | | | |
| RX | MEDLINE-20578749; PubMed-11137999; | | | | |
| RA | Scott H.S., Rudoh J., Wattenhofer M., Shibuya K., Berry A., Chrest R., | | | | |
| RA | Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S., | | | | |
| RA | Youniss F., Mehdi S.Q., Radhakrishna U., Papasavvas M.P., Gehrig C., | | | | |
| RA | Rossier C., Korostilshvsky M., Gal A., Shimizu N., Bonne-Tamir B., | | | | |
| RA | Antonarakis S.E.; | | | | |
| RT | "Insertion of beta-satellite repeats identifies a transmembrane | | | | |
| RT | protease causing both congenital and childhood onset autosomal | | | | |
| RT | recessive deafness". | | | | |
| RL | Nat. Genet. 27:59-63(2001). | | | | |
| CC | - FUNCTION: PROBABLE PROTEASE. | | | | |
| CC | - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL). | | | | |
| CC | - ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/C, D AND | | | | |
| CC | TRUNCATED/TADG-12V; ARE PRODUCED BY ALTERNATIVE SPLICING. | | | | |
| CC | - TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. THE TRUNCATED | | | | |
| CC | ISOFORM IS FOUND AT INCREASED LEVELS IN SOME CARCINOMAS. | | | | |
| CC | - DISEASE: DEFECTS IN TPRESS3 ARE A CAUSE OF TWO FORMS OF AUTOSOMAL | | | | |
| CC | NEUROSENSORY CHILDHOOD-ONSET FORMS OF DEAFNESS, DFNB8 AND DFNB10. | | | | |
| CC | - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE | | | | |
| CC | TRYPSIN FAMILY. | | | | |
| CC | - SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN. | | | | |
| CC | - This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (see http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@isb-sib.ch). | | | | |

9/27/2001

DR EMBL: AF201380; AAC37012.1; -;
 DR EMBL: AB038157; BAB20077.1; -;
 DR EMBL: AB038158; BAB20078.1; -;
 DR EMBL: AB038159; BAB20079.1; -;
 DR EMBL: AB038160; BAB20080.1; -;
 DR MIM: 605511; -;
 DR MIM: 601072; -;
 DR MIM: 605316; -;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LDL_recept_A.
 DR InterPro: IPR001234; trypsin.
 DR Pfam: PF00057; ldl_recept_a; 1.
 DR Pfam: PF00530; SRCR; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR SMART: SM00192; ldl_a; 1.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00202; tryp_spec; 1.
 DR PROSITE: PS01209; LDLRA_1; 1.
 DR PROSITE: PS50068; LDLRA_2; 1.
 DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS50287; SRCR_2; 1.
 DR PROSITE: PS50240; TRYPsin_DOM; 1.
 DR PROSITE: PS00134; TRYPsin_HIS; 1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 KW Hydroxylase; Serine protease; Transmembrane; Signal-anchor; Deafness;
 KW Alternative splicing; Polymorphism.
 FT DOMAIN 1 48
 FT TRANSEM 49 69
 FT DOMAIN 70 454
 FT DOMAIN 72 108
 FT DOMAIN 109 205
 FT DOMAIN 217 454
 FT ACT_SITE 257 257
 FT ACT_SITE 304 304
 FT ACT_SITE 401 401
 FT SITE 216 217
 FT DISULFID 73 85
 FT DISULFID 79 98
 FT DISULFID 92 107
 FT DISULFID 129 194
 FT DISULFID 142 204
 FT DISULFID 207 324
 FT DISULFID 242 258
 FT DISULFID 338 407
 FT DISULFID 370 386
 FT DISULFID 397 425
 FT CARBOHYD 221 221
 FT VARSPPLIC 1 127
 FT VARSPPLIC 318 454
 FT VARSPPLIC 261 293
 FT VARSPPLIC 294 454
 FT VARIANT 53 53
 FT CONFLICT 46 54
 FT CONFLICT 90 90
 FT CONFLICT 350 350
 FT CONFLICT 369 395
 FT CONFLICT 427 427
 FT CONFLICT 454 AA; 49404 MW; 57ECC3678FDD6A6F CRC64;
 SO SEQUENCE

Query Match 100.0%; Score 2447; DB 1; Length 454;
 Best Local Similarity 100.0%; Pred. No. 6; 9e-201;
 Matches 454; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGENDPRAVEAPSEFSLFGIDLDKISVPADDAVAQAQILSLPLKFPPIIYIGIILI 60
 Db 1 MGENDPRAVEAPSEFSLFGIDLDKISVPADDAVAQAQILSLPLKFPPIIYIGIILI 60
 QY 61 LALAIGIHFDSCGKYRCSSFEKIELIARCDGVSDCKGDEDEYRCVRVGGNAVLQVE 120
 Db 61 LALAIGIHFDSCGKYRCSSFEKIELIARCDGVSDCKGDEDEYRCVRVGGNAVLQVE 120
 QY 121 TAASWKTMCDDMKGHANANACQQLGPPSYSSNLRVSSLEQFRFEFSIDHLLPDDK 180
 Db 121 TAASWKTMCDDMKGHANANACQQLGPPSYSSNLRVSSLEQFRFEFSIDHLLPDDK 180
 QY 181 VTALHSHVYREGCASHVVTLOCTACGHRGYSRRIVGNSMLLSQMPQASLOFGYH 240
 Db 181 VTALHSHVYREGCASHVVTLOCTACGHRGYSRRIVGNSMLLSQMPQASLOFGYH 240
 QY 241 LCGSVITPIMITTAHCYVDLYLPKSMTOVGLVSLDNPAPSHLYEKITVHSKYPKR 300
 Db 241 LCGSVITPIMITTAHCYVDLYLPKSMTOVGLVSLDNPAPSHLYEKITVHSKYPKR 300
 QY 301 LGNDIALMKLAGPLTFENMIQPCLPNSENFPDGKCYKTSNGATEEDGAGDASPVLNHA 360
 Db 301 LGNDIALMKLAGPLTFENMIQPCLPNSENFPDGKCYKTSNGATEEDGAGDASPVLNHA 360
 QY 361 AVPLISKICNHRDVGIIISPSMLCAGYLTGGVSDCGSGGPLYQOEERLKLVGATS 420
 Db 361 AVPLISKICNHRDVGIIISPSMLCAGYLTGGVSDCGSGGPLYQOEERLKLVGATS 420
 QY 421 FGIGCAEVNRPQGYTVRTSFLDWIHQEMERDLKT 454
 Db 421 FGIGCAEVNRPQGYTVRTSFLDWIHQEMERDLKT 454
 RESULT 2
 TMS2_MOUSE STANDARD; PRT; 490 AA.
 ID TMS2_MOUSE
 AC 09JIO8; 09JIK4; 09QY82;
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-) (EPITHELIALIN) (PLASMIC
 DE TRANSMEMBRANE PROTEIN X).
 GN TMPRSS2.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC STRAIN=BALB/C;
 RX MEDLINE=21104370; PubMed=11169526;
 RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viiko P.T.;
 RT "Expression of transmembrane serine protease TMPSR2 in mouse and
 human tissues";
 RL J. Pathol. 193:134-140(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Han J., Kim S.;
 RT "Putative transmembrane protease X";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
 RT "A novel mosaic serine protease, epithelialin";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN PROSTATE AND KIDNEY.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: AF199362; AAF97867.1; -
DR EMBL: AF243500; AAF64186.1; -
DR EMBL: AF113596; AAF21308.1; -
DR HSSP: P00761; IAKS.
DR MGD: MGI:1354381; Tmpres2.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR001190; SRCR.
DR InterPro: IPR001254; trypsin.
DR Pfam: PF00057; ldl_recept_a; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PRO0722; CHYMOTRYPSIN.
DR SMART: SM00192; LDLa; 1.
DR SMART: SM00202; SR; 1.
DR SMART: SM00202; Tryp_Spc; 1.
DR PROSITE: PS01209; LDLA_1; 1.
DR PROSITE: PS00068; LDLA_2; 1.
DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
DR PROSITE: PS0287; SRCR_2; 1.
DR PROSITE: PS0240; TRYPsin_DOM; 1.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
FT DOMAIN 1 83 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 84 104 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 105 490 (POTENTIAL).
FT DOMAIN 111 149 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 150 242 LDL-RECEPTOR CLASS A.
FT DOMAIN 254 490 SRCR.
FT ACT_SITE 294 294 SERINE PROTEASE.
FT ACT_SITE 343 343 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 439 439 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 253 254 CLEAVAGE (POTENTIAL).
FT DISULFID 76 125 BY SIMILARITY.
FT DISULFID 119 138 BY SIMILARITY.
FT DISULFID 132 147 BY SIMILARITY.
FT DISULFID 171 230 BY SIMILARITY.
FT DISULFID 164 240 BY SIMILARITY.
FT DISULFID 243 363 BY SIMILARITY.
FT DISULFID 279 295 BY SIMILARITY.
FT DISULFID 408 424 BY SIMILARITY.
FT DISULFID 435 463 BY SIMILARITY.
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 122 122 S -> L (IN REF. 3).
FT CONFLICT 178 178 S -> N (IN REF. 3).
FT CONFLICT 320 320 Y -> H (IN REF. 1).
FT CONFLICT 474 474 N -> D (IN REF. 1).
SO SEQUENCE 490 AA; 53479 MW; 07D2B03EAD8A1A9 CRC64;
Query Match 35.4%; Score 865.5; DB 1; Length 490;
Best Local Similarity 40.9%; Pred. No. 3.4e-66;
Matches 193; Conservative 66; Mismatches 146; Indels 67; Gaps 14;
OY 28 PVAPADANAQAQLSLPLKFPPIIV-----IGII-----A 58
DB 31 PVAPNG-----YNLYPAQYVPSVPOYAPRITTOASTSVIHHTPKSSGAPCTSKSKS 83
OY 59 LILALALIGI-----HFDCS-GKYRCRSSFKCELILARDDGSDCKDGED 103
DB 84 LCLALALGIVLIGAAVAVALMRFWDNSCTSEMECGSGGCTCISSLSLWCDGVAHCPCNGED 143

```

```

OY 104 EYKCVBVGQMAVLQVETA--ASWKTMCSDDMKGHYANVACAOLGPP-SYVESDNLRVSS 160
DB 144 ENRCVAVLYQSFLLQYSSQORRAMYVQCDDMSSESGRAACADMGNKNFYSSQGIIPDS 203
OY 161 LECQFREEFVSIDHLLPDDKVYALHHSYVREGCAGHVVTLOCTACGHR- GYSSRIYV 219
DB 204 GATSPFKLVNNSGVN---DLYKKLYHS---DSCSRMVVSLRJCIBGVRSVRSQRIYV 256
OY 220 GNMSLISQPMQASLOPQGYHLCGGSVITPMTITTAHCVYD-LYLPKSMITQVGLV--S 276
DB 257 GLNASPDGMPWQVSLHVOGVHVGSGSIITPEWTVTAHCVBEPLSGPRYTAFAGITLROS 316
OY 277 LDNPAPSHLVKIVYHSYKPKRLGNDPALKKLACPLNFNEIOPVCJPNSEBENPDCK 336
DB 317 LMFY-GSRIVQEVKIVHPYNDSTKKNDAKLQPLFLFNDLYKVCILPFGAMLDLQ 375
OY 337 VCVTSGWATGDCGADGAPVNLHAAVPLISNKNHSDYVGGITSPSLCAGYLTVGVDS 396
DB 376 ECWISGCGATVE-KGRTSDVLANAAMPLEPSKCNKSYIYNLITPAMICAGFLQGVDS 434
OY 397 CGDSCGPLVCOERRLMKLVGATSEIGCAEVNKPQVYTRVYSFLDMIHQM 448
DB 435 CGDSCGPLVTLKNGITWLIIGDTSWGSCAKALRPGVYGNVTFIDWIVQM 486
RESULT 3
TMS2_HUMAN STANDARD; PRT; 492 AA.
ID TMS2_HUMAN
AC O15393;
DT 15-JUL-1998 (Rel. 36, Created)
DR 15-JUL-1998 (Rel. 36, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-).
GN TMRSS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97468144; PubMed=9325052;
RA Paoloni-Giacobino A., Chen H., Peltsch M.C., Rossier C.,
RA Antonarakis S.E.;
RT "Cloning of the TMRSS2 gene, which encodes a novel serine protease
RT with transmembrane, LDLra, and SRCR domains and maps to 21q22.3.";
RL Genomics 44:309-320(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed=11414763;
RA Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;
RT "Mutation analyses of 268 candidate genes in human tumor cell lines.";
RL Genomics 74:352-364(2001).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease TMRSS2 in mouse and
RT human tissues.";
RL J. Pathol. 193:134-140(2001).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO
CC EXPRESSED IN PROSTATE, COLON, STOMACH, AND SALIVARY GLAND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


FT DOMAIN 71 455 (POTENTIAL).
 FT DOMAIN 112 207 EXTRACELLULAR (POTENTIAL).
 FT ACT_SITE 258 455 SRCR.
 FT ACT_SITE 258 455 SERINE PROTEASE.
 FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 217 218 CLEAVAGE (POTENTIAL).
 FT DISULFID 135 196 BY SIMILARITY.
 FT DISULFID 148 206 BY SIMILARITY.
 FT DISULFID 209 328 BY SIMILARITY.
 FT DISULFID 243 259 BY SIMILARITY.
 FT DISULFID 374 390 BY SIMILARITY.
 FT DISULFID 401 429 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 14 MISSING (IN ISOFORM 2).
 FT VARSPLIC 1 10 MISSING (IN ISOFORM 3).
 FT VARSPLIC 182 182 MISSING (IN ISOFORM 1).
 FT VARSPLIC 183 192 GGLVEAWKP -> MEAOGILIV (IN ISOFORM 1).
 FT CONFLICT 325 325 D -> G (IN REF 1; BAB20277).
 SQ SEQUENCE 455 AA; 49632 MW; 5FCF31789C6899AA CRC64;

Query Match 28.8%; Score 704; DB 1; Length 455;
 Best Local Similarity 35.1%; Pred. No. 1.8e-52;
 Matches 150; Conservative 63; Mismatches 162; Indels 52; Gaps 11;

QY 52 VIGIALLALAIAG--LGIHEDSGKRYCRSSFKCIELIARCDGVSDCKGEDEYR-- 106
 DB 50 VILGVLGLAGAGIASMLVLYLMPAPSPISIGTLQEEENTLNCPEVS-C---EEELIUS 105
 QY 107 -----CVRRGGONAVYQVTTAA--SKRTWCSDMKGHYANVACAQIGFSPYSSDNLRVS 159
 DB 106 LPTVSFRIINGEDLLQOVVRRAPDWLLVCHGMSPALCMHICKSLGHLRLTQHKAVNLIS 165
 QY 160 SLEQREFEVSIDHLLPDDKYATLHNSVYVREG-----CASGHVVTLOCTAC 207
 DB 166 DILNLSGEPAAQS-----ARRGGLVEAWKPSANCPSPGRIVSLKSCSC 209
 QY 208 GHRGYSSTRIVGNSMLLSQMPWQASLQFQYHLCGGSVITPLMTTAAHCYVDLYLPR- 266
 DB 210 G-ARPLASRIYGVGAOVAASRMPQASVMLGSRHTCGASVLAHPWVYTAACHMSFRLSRL 268
 QY 267 -SWTIQVGLVSLDNPAPSH---LVETIYHSHKYPKRGINDIALMKLAGPLTFNEMIQP 322
 DB 269 SSRRVHAGLVS--HGAVRHOGHTWEKILPHPLYSQNHNDYDALLQRTPLNFSDTYDA 326
 QY 323 VCIIPNSEENFPDGKVCWMTSGWATEDGADASFLNHAAPLISNKCINHRDVGGIISP 382
 DB 327 VCLPAKEQYFPMGSCQWGWGHTDPSHNSDPLQDTWVPLISLTLCKSSCMGSCALNH 386
 QY 383 SMCAGYLTVGVDSCGDSGGPLVCOERLMLKLGATSGIGCAEYVKNKFGVYTRVTSFLD 442
 DB 387 RMLCAGYLDGRADACGDSGGPLVCPGSDTWHLVGVVSWGRCGAEBNRPBGVYAKVAEFLD 446
 QY 443 WHEOME 449
 DB 447 WIDHIVQ 453

RESULT 5
 TMS5 HUMAN STANDARD; PRT: 457 AA.
 AC 09H3S3;
 DT 20-AUG-2001 (rel. 40; Created)
 DT 20-AUG-2001 (rel. 40; Last sequence update)
 DE TRANSMEMBRANE PROTEASE, SERINE 5 (EC 3.4.21.-) (SPINESIN).
 GN TMPRSS5.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Mitsui S., Yamaguchi N.;
 RT "Molecular cloning of human spinosin."
 RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AB028140; BAB20375.1; .
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR InterPro: IPR001314; Chymotrypsin.
 DR Pfam: PF000089; trypsin.1.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00202; TRYP-SPEC; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; FALSE_NEG.
 DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS0287; SRCR_2; FALSE_NEG.
 DR Hydrolyase; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein.
 FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT FT (POTENTIAL).
 FT DOMAIN 71 457 EXTRACELLULAR (POTENTIAL).
 FT FT (POTENTIAL).
 FT DOMAIN 112 207 SRCR.
 FT ACT_SITE 258 457 SERINE PROTEASE.
 FT ACT_SITE 258 457 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 217 218 CLEAVAGE (POTENTIAL).
 FT DISULFID 135 196 BY SIMILARITY.
 FT DISULFID 148 206 BY SIMILARITY.
 FT DISULFID 209 328 BY SIMILARITY.
 FT DISULFID 243 259 BY SIMILARITY.
 FT DISULFID 374 390 BY SIMILARITY.
 FT DISULFID 401 429 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 457 AA; 49574 MW; 64406BA983A2651 CRC64;

Query Match 28.4%; Score 696; DB 1; Length 457;
 Best Local Similarity 35.2%; Pred. No. 8.6e-52;
 Matches 149; Conservative 58; Mismatches 176; Indels 40; Gaps 7;

QY 53 VIGIALLALAIAGLGIHEDSGKRYCRSSFKCIELIARCDGVSDCKGEDEYRCVRYVG 112
 DB 51 VILGVLGLAGAGVGSMLV---LYLCPAASQPI-----GLQDEEITLSCSEASA 98
 QY 113 QNAVY-----QVFTASWKTMCSDMKGHYANVACAQIGFSPYVS 152
 DB 99 EEWLLPALKVTVSRINSEDFLENAQVDPRLVLCHEGMSPALGIDICWSLGLRLTH 158

RESULT 7
ENTR MOUSE STANDARD: PRT: 1069 AA.

AC P97435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
GN PRS57 OR ENTRX
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Ductum;
RX MEDLINE=98147142; PubMed=946188;
RA Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;
RT "Structure of murine enterokinase (enteropeptidase) and expression in
small intestine during development.";
RT Am. J. Physiol. 274.G342-G349(1998).
CC - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
CC - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-|-ILE-7 BOND IN
TRYPSINOGEN.
CC - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
SIMILARITY).
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
CC - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC - SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U73378; AAB37317.1; -
DR HSSP: P00763; IDPO.
DR MEROPS: S01.156; -
DR MGD: MGI:1197523; PRS57.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR000998; MAM.
DR InterPro: IPR000082; SEA.
DR InterPro: IPR001190; SRCR.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00057; ldl_recept_a; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF01390; SEA; 1.
DR Pfam: PF00530; SRCR; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00192; LDLa; 2.
DR SMART: SM00137; MAM; 1.
DR SMART: SM00200; SEA; 1.

DR SMART: SM00202; SR; 1.
DR SMART: SM00202; TRYP_Spc; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01209; LDLa_1; 2.
DR PROSITE: PS01209; LDLa_2; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS00740; MAM_2; 1.
DR PROSITE: PS50060; MAM_2; 1.
DR PROSITE: PS50024; SEA; 1.
DR PROSITE: PS50028; SRCR_1; FALSE_NEG.
DR PROSITE: PS50028; SRCR_2; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HTS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Signal-anchor; Glycoprotein; Myristate; Hydrolase;
KW Serine protease; Zymogen; Transmembrane; Repeat;
NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT CHAIN 1 829
FT TRANSMEM 19 1069
FT DOMAIN 52 169
FT DOMAIN 227 268
FT DOMAIN 270 379
FT DOMAIN 387 549
FT DOMAIN 569 679
FT DOMAIN 686 724
FT DOMAIN 723 816
FT DOMAIN 830 1069
FT ACT_SITE 874 874
FT ACT_SITE 925 925
FT ACT_SITE 1021 1021
FT LIPID 2 2
FT DISULFID 229 242
FT DISULFID 236 255
FT DISULFID 249 266
FT DISULFID 688 700
FT DISULFID 695 713
FT DISULFID 707 722
FT DISULFID 817 875
FT DISULFID 859 1027
FT DISULFID 959 1006
FT DISULFID 991 1006
FT CARBOHYD 1017 1045
FT CARBOHYD 147 147
FT CARBOHYD 197 197
FT CARBOHYD 212 212
FT CARBOHYD 373 373
FT CARBOHYD 380 380
FT CARBOHYD 433 433
FT CARBOHYD 515 515
FT CARBOHYD 579 579
FT CARBOHYD 675 675
FT CARBOHYD 727 727
FT CARBOHYD 751 751
FT CARBOHYD 770 770
FT CARBOHYD 791 791
FT CARBOHYD 897 897
FT CARBOHYD 936 936
FT CARBOHYD 999 999
SQ SEQUENCE 1069 AA; 118735 MW; E62549E463743C3D CRC64;

Query Match 27.0%; Score 661.5; DB 1; Length 1069;
Best Local Similarity 37.6%; Pred. No. 2e-48;
Matches 152; Conservative 64; Mismatches 143; Indels 45; Gaps 16;
OY 67 LGIHFDG-SGKYRCSFPCIELIARCDSVSDCKDEDEYRCVY---VGQNAVQLQVFT 121
DB 682 LGIPEPCQDEDFQCKDG-KCIPLGNCDSYRHCRCDSDEASCVRFNGTRNSNGLVQFNI 740
OY 122 AASMKTMCSDDMKGHYANACAGDLGFPYSSNINLAVSISLEGFRREYPSIDHLLDDDKY 181
DB 741 HSIWHTACAEWMTTQISNEVCHLLDGS--ANSMPISSTGG--GPFVRVND----- 788
OY 182 TALHSVYVREG--CASGHVVTLOCT--ACGHR--GYSSRIYVGNMISLQWPMQASL 234

```

Db 789 -APNGSLILPISLQCSODSILLQCNKSCGCKVKVTKVSPKIVGSGDAGAMPVVAL 847
QY 235 QFGCYH-----LCGGSVITPMTITAHACYDYL-ERSWTOYGL--VSLIDNP-A 282
Db 848 ----YHNRDSTDRLLCGALVSSDMLVAHCVRRNLDPRTMVAJGLHMOSLSTPQV 903
QY 283 PSHIVKIVYHSKYPKRLGNDIALMKLAPLEFENEMIOPLCPNSEENFPDGKVCWTSG 342
Db 904 VRRVVDIVINPHDRKRRKNNDIAMHLEKRVNTDIOPLCPLEENQIFIPGTCIAG 963
QY 343 WGATEDGAGASPYLNAHVAAPLISNKTCHNR-DVYGGIISPSMLCAGYLTCGVSCGDS 401
Db 964 WGYDKINAGSTVDVLEKADVDPLISNEKCOOOLPEYN--ITESHICAGYEKGIDSCGDS 1021
QY 402 GGPLYCOERLKLKLVGATSPGICAEVKNKGVYTRVTSFLDWIH 445
Db 1022 GGPLMCENNKMFVLVGTISFGVOCALPNHPGVYRVSOFLTEWIH 1065

RESULT 8
TMS4_HUMAN STANDARD: PRT: 437 AA.
ID TMS4_HUMAN STANDARD: PRT: 437 AA.
AC Q9NRK4; Q9NZAS;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE TRANSMEMBRANE PROTEASE, SERINE 4 (EC 3.4.21.-) (MEMBRANE-TYPE SERINE
DE PROTEASE 2) (MT-SP2).
GN TMRSS4 OR TMRSS3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic carcinoma;
RX MEDLINE=20283276; PubMed=10825129;
RA Walltrap C., Haehnel S., Mueller-Pillasch F., Burghardt B.,
RA Iwanura T., Rubenbueger M., Lerch M.M., Adler G., Gress T.M.;
RT "A novel transmembrane serine protease (TMRSS3) overexpressed in
RT pancreatic cancer.";
RL Cancer Res. 60:2602-2606(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Sneegeens S.S., Lorimer D.D., Wang E., Hou J., Llanevers C.;
RT "MT-SP2, a novel type II membrane serine protease expressed in
RT trachea, colon, and small intestine: identification, cloning, and
RT chromosomal localization.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PLAY A ROLE IN TUMOR PROGRESSION, METASTASIS
CC FORMATION AND TUMOR INVASION.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN PANCREATIC, GASTRIC, COLORECTAL
CC AND AMPULLARY CANCER. VERY WEAK EXPRESSION IN NORMAL
CC GASTROINTESTINAL AND UROGENITAL TRACT.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF179224; AAF74526.1; -
CC EMBL; AF216312; AAF31436.1; -
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR002172; LDL_receptL_A.

```

```

DR InterPro: IPR001190; SRCR.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF000057; ldl_receptL_a.1.
DR Pfam: PF000089; trypsin.1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDLA.1.
DR SMART: SM00202; SR.1.
DR SMART: SM00202; TRYP_SPE.1.
DR PROSITE: PS01209; LDLA.1; FALSE_NEG.
DR PROSITE: PS00668; LDLA.2; FALSE_NEG.
DR PROSITE: PS00420; SRCR.1; FALSE_NEG.
DR PROSITE: PS50287; SRCR.2; 1.
DR PROSITE: PS50240; TRYPSIN_DOM.1.
DR PROSITE: PS00134; TRYPSIN_HIS.1.
DR PROSITE: PS00135; TRYPSIN_SER.1.
KW Hydroxylase, Serine protease; Transmembrane; Signal-anchor.
FT DOMAIN 1 38
FT TRANSMEM 1 38
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT DOMAIN 60 437
FT DOMAIN 61 93
FT DOMAIN 94 194
FT DOMAIN 205 437
FT ACT_SITE 245 245 SERINE
FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 204 205 CLEAVAGE (POTENTIAL).
FT DISULFID 64 83 BY SIMILARITY.
FT DISULFID 77 92 BY SIMILARITY.
FT DISULFID 127 183 BY SIMILARITY.
FT DISULFID 140 193 BY SIMILARITY.
FT DISULFID 196 310 BY SIMILARITY.
FT DISULFID 230 246 BY SIMILARITY.
FT DISULFID 356 372 BY SIMILARITY.
FT DISULFID 383 410 BY SIMILARITY.
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1 31 MLODPSDQPLNSLDVKKPRIPMETFRK -> MSNPCA
FT NPVSPKPSRS (IN REF. 2).
SQ SEQUENCE 437 AA; 48204 MW; 351B2FD4A8657B12 CRC64;

Query Match 27.0%; Score 660.5; DB 1; Length 437;
Best Local Similarity 35.4%; Pred. No. 8.6e-49;
Matches 157; Conservative 74; Mismatches 149; Indels 63; Gaps 17;

QY 44 LPLKFFIYIGTALLALA-----IGGHPDCGKRCRSSPFC--IELIAR--- 91
Db 24 IPMETFRKVGIPITIALLSLIIIVVLIKVLID--KT----YELCGPLHPIPKOL 76
QY 92 CDGVSDEKDEDEYRC-----VRVGONAVLQVTTAA--SWKTMCSDDMKGHYA 138
Db 77 CDGLDLPDEDEDEHCKKSPFEGNAVAVRLSKRSTQIVLDSATGNMFSACPFNPTALA 136
QY 139 NVACAQIGPSPYSSONLRAVSSLEGQFEEFVSTIDLLPDKVTALHHSYVREG---CA 195
Db 137 ETACRQMGYS-----KPTFRAYEISPDQDLDAVEITENSQELRMNRSSGCL 184
QY 196 SGHVTYTCACGHRGYSRIYIGNMSILSOMPQASLOFGYHLCGGSVTPMTITA 255
Db 185 SGSLVSLHCLACG-KSKLTPRVVGGEEASVDSMWOVSIOYDKOHVCGGSLIDPHWVLA 243
QY 256 AHCV--YDLYLPKSWTIQVGLSLDNPAPSHLVKIV---YHSKYPKRLGNDIALMK 309
Db 244 AHCRKRTDVF---NMKVRAGSDKL--GSPPLSLAVAKIIIEENPMV-PR--DNIDIALMK 295
QY 310 LAGPLTFENEMIOPLCPNSEENFPDGKVCWTSGMGATEDGAGASPYLNAHVAAPLISNKT 369
Db 296 LQPLTFSGTVRPICLPFEEDELTPTATPLWITIGWGTFGKNGKMSDILLQASVOVIDSTR 355
QY 370 CNHRDVGGIISPMCLAGLTGGVSCGDSGGPLVCOGRRRLMKLVGATSPGICAEVN 429
Db 356 CNMDDAYOGVEYTERKMCAGIPEGGVDTCCGDSGGLPMYOSDO--MHVVGIVSWGTCGGPS 414

```


| FT | CARBOHYD | 179 | 179 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
|-----------|---|-----------------------------------|--------------------|--|--------------|
| FT | CARBOHYD | 328 | 328 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 335 | 335 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 388 | 388 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 440 | 440 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 470 | 470 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 503 | 503 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 534 | 534 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 630 | 630 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 682 | 682 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 706 | 706 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 725 | 725 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 848 | 848 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 887 | 887 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 909 | 909 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 949 | 949 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CONFLICT | 134 | 134 | O -> E (IN REF. 3). | (POTENTIAL). |
| FT | CONFLICT | 732 | 732 | S -> P (IN REF. 3). | (POTENTIAL). |
| FT | CONFLICT | 754 | 771 | SOCLODLSIRLCQNHKS -> RRNAKNEIDALSPITLIA (IN REF. 3). | (POTENTIAL). |
| FT | SEQUENCE | 1019 AA; | 112923 MW; | B6AA4245F6DA4563 CRC64; | (POTENTIAL). |
| QY | Query Match | 26.8%; | Score 657; | DB 1; | Length 1019; |
| QY | Best Local Similarity | 37.7%; | Pred. No. 4.6e-48; | | |
| QY | Matches 148; | Conservative 67; | Mismatches 148; | Indels 30; | Gaps 15; |
| QY | 67 LGIHFD-C-SGRVRCSSFKCIELARCDGVSDCKGDEYRCVR- - - VGGONAVIQVFT 121 | . . . : : : : : : : . : : : : | | | |
| QY | 637 LGIPRCACADHFQCKNG- ECVPIVNLCDGHLHCEGSDPADCVRFNCTNNNGIVRRI 695 | | | | |
| QY | 122 AASWTKMCSDDMKGHYANVACAQJGFPSYVSSDNLRVSSLEQGFREFEVSIDHLLPDDKV 181 | . . . : : : : : : : . : : : : | | | |
| QY | 696 QSIWHTACAEWMTQIISNDVQCLLGISGNSSK- -PIRSTDG- -GPVVKL-NTAPDG- - 747 | | | | |
| QY | 182 TALHSYVYREGCASHVYVLTQCT- -ACGHR- - -RGYSRIYVGMSLLSQMPQASIQF 236 | . . . : : : : : : : . : : : : | | | |
| QY | 748 - - -HLITPSSQOCLODLSIRLCQNHKSCGKKLAADITPIKIVGGSNAEGAMPVAVGLTY 804 | | | | |
| QY | 237 QGYHLGGSVTPPLIITIAACVVDLYL-PRKSTIQVCL- -VSLIDNP-ARSHLEKIVY 292 | . . . : : : : : : : . : : : : | | | |
| QY | 805 GGRLLCGSLVSSDMLVSAACHVGRNLPEPSKWTALILGLHKHSNLTSPQYPRLLDELYI 864 | | | | |
| QY | 293 HSKYKPKRLGNDIALMKLAGPLTFENEMIQPVCLPNSSENPFDGKVCWTSGCATGEDGAGD 352 | . . . : : : : : : : . : : : : | | | |
| QY | 865 NPHNRRKRDIDAMHLEFQVNTDYTIQJCLDEPNQVFPFGRCISJAGMG-FTVYQGT 923 | | | | |
| QY | 353 ASPVLNHAAYVLLSNKTCINHR-DVYGGIISSMILCACTLVGCVSQCQSDSGPLVCOGRR 411 | . . . : : : : : : : . : : : : | | | |
| QY | 924 TANTIQEADVPLLSNERCQOQMPEN-ITENMTCAGYEGBIDSCQDSDGGPLMCOENN 981 | | | | |
| QY | 412 LMKLVGATSFSGICAEVKNPQVYTRVTSFLMI 444 | . . . : : : : : : : . : : : : | | | |
| QY | 982 RMFLAGVTSFGYKCALPRRPGVYARVSKFTEMI 1014 | | | | |
| RESULT 11 | ENTRK_PIG | STANDARD; | PRT; | 1034 AA. | |
| AC | ENTRK_PIG | STANDARD; | PRT; | 1034 AA. | |
| AC | P98074; | | | | |
| DT | 01-FEB-1996 (Rel. 33, Created) | | | | |
| DT | 01-FEB-1996 (Rel. 33, Last sequence update) | | | | |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) | | | | |
| DE | ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKININASE). | | | | |
| GN | PREC7 OR ENTK. | | | | |
| OS | Sus scrofa (Pig). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus. | | | | |
| OX | NCBI_TaxID=9823; | | | | |
| OX | [1] | | | | |
| RP | SEQUENCE FROM N.A., AND PARTIAL SEQUENCE. | | | | |
| RC | TISSUE=Duodenal mucosa; | | | | |
| KX | MEDLINE=94327546; PubMed=8051081; | | | | |

QY 351 GASPVLNHAAPLISNKCINHR-DVYGGIISPSMLCAGLTGGVDSGCGSGPLVCOE 409
 DB 937 GSPADILGADVPLLSNEKCCQOMPEYN--ITENMCMACGEEGIDSCGDSGPIMLCE 994
 QY 410 RLRLKLVGATSFSGICAEVKKPGVYTRVTSFLDWT 444
 DB 995 NNRMLAGVTSFGYOCALPRRPGVAVRPKFTEMI 1029

RESULT 12
 HEP5_MOUSE STANDARD: PRT: 416 AA.
 AC 035453;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE SERINE PROTEASE HEP5IN (EC 3.4.21.-).
 GN HPN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=98058912; PubMed=9395459;
 RA Yu T.-H., Liu R.-W., Haakma C., Tomasek J.J., Howard E.W.;
 RT Identification and cloning of the membrane-associated serine
 protease, hepsin, from mouse preimplantation embryos.;
 RL J. Biol. Chem. 272:31315-31320(1997).
 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
 OF CELL MORPHOLOGY.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

DR EMBL: AF030065; AAB84221.1; -
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.224; -
 DR MGD: MGI:1196620; Hpn.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00202; SR.1.
 DR SMART: SM00202; TRYP-SPEC.1.
 DR PROSITE: PS00240; TRYP-SIN.DOM.1.
 DR PROSITE: PS00134; TRYP-SIN.HIS.1.
 DR PROSITE: PS00135; TRYP-SIN.SER.1.
 KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
 FT CHAIN 1 161 NON-CATALYTIC CHAIN (POTENTIAL).
 FT CHAIN 162 416 CATALYTIC CHAIN (POTENTIAL).
 FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 17 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 (POTENTIAL).
 FT DOMAIN 44 416 EXTRACELLULAR (POTENTIAL).
 FT ACT_SITE 162 416 SERINE PROTEASE.
 FT ACT_SITE 202 416 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 256 416 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 352 416 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 152 276 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 187 203 BY SIMILARITY.
 FT DISULFID 321 337 BY SIMILARITY.
 FT DISULFID 348 380 BY SIMILARITY.

FT CARBOHYD 111 111 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 416 AA; 44739 MW; 4321944FE4004FB48 CRC64;
 Query Match 26.6%; Score 650; DB 1; Length 416;
 Best Local Similarity 38.5%; Pred. No. 6,3e-48;
 Matches 136; Conservative 49; Mismatches 128; Indels 40; Gaps 8;

QY 121 TASMKTCSDDMKHYANACAGLFPSPVSSDNLRYSLDEQFREFPSID----- 173
 DB 68 TEGTWRLCSSRSNARVAGLGECEMGLRALASELDVTRAGANGTSGFCEVDEGLPLA 127
 QY 174 -HLPPDKVYALHHSYVREGCAGHYVTLQCTACGHRGSSRIYGGNNLSLQMPWQA 232
 DB 128 QRL- DVLVVC-----DCPRGRFLATCCDGRKRLPVRIYGGQSSLGRRMPWQY 177
 QY 233 SLQFGYHLGGSVTPPLMTITRAHCYVDLYLPE-----SWTIQYGLVSLDNPAPSHL 286
 DB 178 SLRYDGTSLCGGSLTSGDWVLTAAHC-----FPERRRVLSRMRVFPAGAVARTSPHAVQLG 232
 QY 287 VEKIVYHSKYKPKR-----LGNDIALMKLAGPLTFENEMTOPVCLPNSSENPFGKVCWT 340
 DB 233 VQAVYTHGGTLPDPPTIDBNSNDIALVHLSSLPLETYIQVCLPAACQALVDGAVCYV 292
 QY 341 SGWATEDGADGASPVLNHAAPLISNKCINHRDVYGGIISPSMLCAGLTGGVDSGCGD 400
 DB 293 TGGNGNQ-FYGGQAMVLOEAPVPIISNEVCNSDFGNGIKRPMFCAGPEGIDACQGD 351
 QY 401 SGGLVCOE---RLRLKLVGATSFSGICAEVKKPGVYTRVTSFLDWT 449
 DB 352 SGGLVCOE---RLRLKLVGATSFSGICAEVKKPGVYTRVTSFLDWT 449

RESULT 13
 HEP5_RAT STANDARD: PRT: 416 AA.
 AC 005511;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SERINE PROTEASE HEP5IN (EC 3.4.21.-).
 GN HPN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=93305733; PubMed=8318546;
 RA Farley D., Raymond F., Nick H.;
 RT Cloning and sequence analysis of rat hepsin, a cell surface serine
 protease.;
 RL Biochim. Biophys. Acta 1173:350-352(1993).
 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
 OF CELL MORPHOLOGY.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

DR EMBL: X70900; CAA50256.1; -
 DR PIR: S32013; S32013.
 DR PIR: S33777; S33777.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.224; -.

DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00240; TRYP_SPE; 1.
 DR PROSITE: PS00240; TRYPsin_DOM; 1.
 DR PROSITE: PS00134; TRYPsin_HIS; 1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 KM Hydroxylase; Serine protease; Transmembrane; Signal-anchor.
 FT CHAIN 1 161
 FT CHAIN 162 416
 FT DOMAIN 1 16
 FT TRANSMEM 17 43
 FT DOMAIN 44 416
 FT ACT_SITE 162 416
 FT ACT_SITE 202 202
 FT ACT_SITE 256 256
 FT ACT_SITE 352 352
 FT DISULFID 152 276
 FT DISULFID 187 203
 FT DISULFID 321 337
 FT DISULFID 348 380
 FT CARBOHYD 111 111
 SQ SEQUENCE 416 AA; 44926 MW; E5A9F8FA9550E180 CRC64;

Query Match 26.4%; Score 645; DB 1; Length 416;
 Best Local Similarity 35.2%; Pred.No. 1.7e-47;
 Matches 146; Conservative 61; Mismatches 156; Indels 52; Gaps 11;

OY 60 ILALAIGLIHFGSCGKRRSSFKCIELIARCDGVSDCKDEGDEKVCVAVGONAVIAY 119
 DB 17 VALVYGTLLFLTGIG---AASMAIVTILR-----SDQPLVYQVQLSPEDSLRLV 65
 OY 120 -FTAASMKTVCSDMKGVHNAVACAOLEFPYSVSDNLVSSLEGQFREEFVSID----- 173
 DB 66 DKTEGTWRLLCSSRSNARVAGLGCCEGEGFLALHSELVDRTAGANGTSFPCVDEGLP 125
 OY 174 ---HLIPDOKVTALHHSYVREGCASGHVYTLCTACGHRGSSRYVGNMSSLQMPW 230
 DB 126 LAQRL--DVISVC-----DCKRFRFLTACQDGRKRLPVDRVGGQSSSLGKMPW 175
 OY 231 QASLOFGYHLGCGSVYTPMLITTAACVYDLVLPK-----SWTIQVGLVSLDNPAPS 284
 DB 176 QVSLARDGTHLGGSGSLSGMWVLTAAHC-----FPERNRRLSRMRVFAAGAVATSPHAYQ 230
 OY 285 HIVEKIVYHSKYPKR-----LNDIATLMKLAGPLTENMIQPVCLPNSSENFDPDGKVC 338
 DB 231 LGVOAVIYHGGYLPFRDPTIDENSNDIALVHLSLPLTEYIQPVCLPAAGQALVGGKVC 290
 OY 339 WISGKATGEGAGDASPVLNHAHVPLISNKHNDVYGGIISPSMCAQYLGVGDSQC 398
 DB 291 TVTGMKNTQ-FYGGQAVVIOEARVPLISNVCSPDYQKIRPKMFCAYPRGGIDACQ 349
 OY 399 GDSGGLVCOER-----RLMKLVGATSEFGICAEVKNKPGVTVRTSFLDWIHEOME 449
 DB 350 GDSGGLVCOER-----RLMKLVGATSEFGICAEVKNKPGVTVRTSFLDWIHEOME 449
 RESULT 14
 KAL_RAT
 ID KAL_RAT STANDARD; PRT; 638 AA.
 AC P14272;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 20-NOV-2001 (Rel. 40, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN PK.
 OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91129236; PubMed=1993180;
 RA Beaudien G., Rosinski-Chupin I., Mattei M.-G., Mbikay M., Chretien M.,
 RT "Gene structure and chromosomal localization of plasma kallikrein.";
 RL Biochemistry 30:1628-1635(1991).
 RN [2]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=90091743; PubMed=2598771;
 RA Seidah N.G., Lidenheim R., Mbikay M., Hamelin J., Lutfalla G.,
 RA Rougeon F., Lazure C., Chretien M.;
 RT "The cDNA structure of rat plasma kallikrein.";
 RL DNA 8:563-574(1989).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
 CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XII, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).
 CC
 DR EMBL; M62357; AAA74563.1; -;
 DR EMBL; M62358; AAA74563.1; JOINED.
 DR EMBL; M62346; AAA74563.1; JOINED.
 DR EMBL; M62349; AAA74563.1; JOINED.
 DR EMBL; M62350; AAA74563.1; JOINED.
 DR EMBL; M62351; AAA74563.1; JOINED.
 DR EMBL; M62352; AAA74563.1; JOINED.
 DR EMBL; M62353; AAA74563.1; JOINED.
 DR EMBL; M62354; AAA74563.1; JOINED.
 DR EMBL; M62355; AAA74563.1; JOINED.
 DR EMBL; M62356; AAA74563.1; JOINED.
 DR EMBL; M30282; AAA41463.1; -;
 DR EMBL; M58590; AAA42069.1; -;
 DR PIR; A39180; KORTPL.
 DR HSSP; P00750; IRTF.
 DR MEROPS; S01.212; -;
 DR InterPro: IPR000177; Apple.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR003014; PAN.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam; PF00024; PAN; 4.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00005; APPLEDOMAIN.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00223; APPLE; 4.
 DR SMART; SM00202; TRYP_SPE; 1.
 DR PROSITE; PS00495; APPLE; 4.
 DR PROSITE; PS00240; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KM Hydroxylase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
 KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
 FT Repeat.
 FT SIGNAL 1 19

```

FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 SERINE PROTEASE.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
FT DISULFID 21 104 BY SIMILARITY.
FT DISULFID 47 77 BY SIMILARITY.
FT DISULFID 51 57 BY SIMILARITY.
FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 345 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
SQ SEQUENCE 638 AA: 71273 MW: 454BBB27EBCAF88 CRC64:

```

Query Match 23.3%; Score 570; DB 1; Length 638;
 Best Local Similarity 40.8%; Pred. No. 6.8e-41;
 Matches 116; Conservative 52; Mismatches 88; Indels 28; Gaps 9;

```

QY 189 YREGAGSHVYTL-----OCTAGHRRGYSRRVGGNMSLSQMPQASIQFO---G 238
    | : | | | | : | | | | | | | | | | | | | | | | | | | | | | | |
DB 362 YEAQG-SSGYSRLCKRVSSDCTT---KINARIVGNTSLGEMPMQVSLQVQLVQ 415
    | : | | | | : | | | | | | | | | | | | | | | | | | | | | | | |
QY 239 YHCGSVTPITTAHCVYDLYLPKSMITQVGLVSL--LDNPAHSLVEKIVYHSY 296
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 416 NIMCGSITIGRMIITLAAHCFGIPDPDWRIYGIILNSETINKTPSSIKELIHQY 475
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 KKRRLGNDIALMKLAGPLTFNEMIQPVCLPNSSENFPPDKVCMTSGWATGAGDASPV 356
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 476 KMEGSDYDIALKLOPLPLNTEFQKPICLPSKADTNTITNCGWVIGWYTKR-REGTONI 534
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 357 LHMHAAPLISAKIC--NHNDVYGGIISPSMLCAGYITGVDSQCGDSGGLVCQERRLMK 414
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 535 LQKATIPLPVNEECCKKYDY---VITKQMICAGYKEGIDACKGDSGLVCKHSGRMQ 591
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 415 LVGGSFSGICAEVKNPGVYTVTSFLDIHROM---ERDQKT 454
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 592 LVGITSWEGCAKREOPGYITVAETIDWILERIQSSKERALET 635
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 15
KAL_MOUSE STANDARD; PRT; 638 AA.
AC P26262;
DT 01-MAY-1992 (rel. 22, Created)
DT 01-MAY-1992 (rel. 22, Last sequence update)
DT 20-UG-2001 (rel. 40, Last annotation update)
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
DE (KININGENIN) (FLETCHER FACTOR).
GN KUK3 OR PK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; TISSUE=Liver;
RX MEDLINE=91090844; PubMed=2264928;
RA Seidaph N.G., Sawyer N., Hamelin J., Mion P., Beaubien G.,
RA Brachapa L., Rochemont J., Mbikay M., Chretien M.;
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
RT and comparison of protein and mRNA levels among species.";
RL DNA Cell Biol. 9:737-748(1990).
CC -I- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -I- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -I- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M58588; AAA6393.1; -
DR PIR: A36557; KOMSPL.
DR HSSP: P00750; IRTF.
DR MEROPS: S01.212; -.
DR MGD: MGI:102849; KIK3.
DR Interpro: IPR000177; Apple.
DR Interpro: IPR001314; Chymotrypsin.
DR Interpro: IPR003014; PAN.
DR Interpro: IPR001254; Trypsin.
DR Pfam: PF00024; PAN; 4.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00005; APPLDOMAIN.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00223; APPLE; 4.
DR SMART: SM00020; TRYP-Spc; 1.
DR PROSITE: PS00495; APPLE; 4.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KM Hydrolyase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT DOMAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 SERINE PROTEASE.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
FT DISULFID 21 104 BY SIMILARITY.
FT DISULFID 47 77 BY SIMILARITY.
FT DISULFID 51 57 BY SIMILARITY.

```

FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 345 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
SQ SEQUENCE 638 AA: 71368 MW: CC27C93AB1086599 CRC64;

Query Match 22.9%; Score 561; DB 1; Length 638;

Best Local Similarity 34.8%; Pred. No. 4e-40;

Matches 126; Conservative 61; Mismatches 109; Indels 66; Gaps 12;

OY 132 DMKGHANACAOQGFPSYSSDNLRVSSLEGGFREFEVSIDHLLPDDKVTALHHSYVR 191
DB 301 DEEGELNV-----TEVOGADVCECTCTKTRCOFF-IYSLRPOD-----CKE 342
OY 192 EGC-----ASGHVYTL-----QCTACGHRGRYSRIYGGNMS 223
DB 343 EGCCKSLRLSTDGSPTRITYGMOGSSGYSLRLCKLVDSPCCT-----KINARIYGGTNA 397
OY 224 LISQWPMQASLOFO---GYHLCGGSVITPLMIITAAHCYDLYLPRKSWTIQVGLVSL-L 278
DB 398 SIGEMPMQVSLQVKIVSQTHLCGSIIGROWVYTAHCFDGIYPDVMRIYGGILSLSEI 457
OY 279 DNPAPSHVEKIVYHSKYPKRLGNDIALMKLAGPLTFNEMIQVCLPNSENFPPDGKVC 338
DB 458 TKETPSSRIKELIHQERYVSEGNVDIALIKLOTPLNTEFOKPLCLPSKADNTNTIYTNC 517
OY 339 WTSQMGATEDGAGDASPVLNHAAPLISNKIC--NHRDVGGIISPSMLCAGYLTGVD 396
DB 518 WYTGMYGTYKE-QGETONILQKATIPLPNBECCQKKRYD---VINKOMICAGYKEGTDA 573
OY 397 CQGDSSGPLVCOERRLMKLVGATSEIGCAEVKKPGVYTRVYSFLDWIHQOME----RDL 452
DB 574 CKGDSGGPLVCKHSGRMQLVGIISWEGCGCRKDQPGVYTKVSEYMDMILEKTQSSDVRAL 633
OY 453 KT 454
DB 634 ET 635

Search completed: April 11, 2002, 08:57:04
Job time: 217 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:52:32 ; Search time 25.01 Seconds
(without alignments)
1344.633 Million cell updates/sec

Title: US-09-846-512-12
2447
Sequence: 1 MGENDPPAVEAPFSEFSLFG.....TRVNSFLDWHEQMEROCLKT 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A.Geneseq_1101.*
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 2432.5 | 99.4 | 453 | 22 | AAE06935 |
| 2 | 2430.5 | 99.3 | 453 | 20 | AAAB4250 |
| 3 | 2418 | 98.8 | 452 | 20 | AAAY41694 |
| 4 | 2375 | 97.1 | 454 | 21 | AAAB32246 |
| 5 | 1788 | 73.1 | 327 | 21 | AAAY72093 |
| 6 | 1354 | 55.3 | 294 | 21 | AAAB32247 |
| 7 | 1339 | 54.7 | 248 | 21 | AAAB3572 |
| 8 | 1034 | 42.3 | 193 | 22 | AAAG73966 |
| 9 | 876.5 | 35.8 | 492 | 21 | AAAY2050 |
| 10 | 876.5 | 35.8 | 492 | 22 | AAAM01315 |
| 11 | 875.5 | 35.8 | 492 | 21 | AAAY77726 |

| | | | | | | |
|----|-------|------|------|----|-----------|---------------------|
| 12 | 875.5 | 35.8 | 492 | 21 | AAAY44406 | Human 20PFI2-GTC2 |
| 13 | 868.5 | 35.5 | 393 | 22 | AAAM01317 | PI000C partial ami |
| 14 | 864.5 | 35.3 | 492 | 21 | AAAB36901 | Human TMPS52 prot |
| 15 | 864.5 | 35.3 | 492 | 21 | AAV57280 | Ovrl15 homolog pro |
| 16 | 864.5 | 35.3 | 492 | 22 | AAE06943 | Human serine prote |
| 17 | 864.5 | 35.3 | 492 | 22 | AAAM01294 | Human transmembran |
| 18 | 731 | 29.9 | 421 | 22 | AAAB5042 | Human SER6 protein |
| 19 | 731 | 29.9 | 421 | 22 | AAE01943 | Human transmembran |
| 20 | 727 | 29.7 | 688 | 22 | AAE01944 | Human transmembran |
| 21 | 704 | 28.8 | 445 | 21 | AAAB1698 | Mouse serine prote |
| 22 | 704 | 28.8 | 445 | 21 | AAAB08950 | Human secreted pro |
| 23 | 702 | 28.7 | 414 | 21 | AAAB08912 | Human secreted pro |
| 24 | 696 | 28.4 | 457 | 21 | AAAB1699 | Human serine prote |
| 25 | 686.5 | 28.1 | 283 | 21 | AAAB1492 | Human prostate-ass. |
| 26 | 667 | 27.3 | 423 | 22 | AAE06944 | Human transmembran |
| 27 | 664.5 | 27.2 | 417 | 22 | AAE06942 | Human enterokinase |
| 28 | 661 | 27.0 | 273 | 21 | AAAB1696 | Mouse serine prote |
| 29 | 660.5 | 27.0 | 435 | 22 | AAV06437 | Human protease HUP |
| 30 | 660.5 | 27.0 | 435 | 22 | AAV72558 | Human seripancrin |
| 31 | 660 | 27.0 | 798 | 15 | AAAB57283 | Bovine enterokinase |
| 32 | 657.5 | 26.9 | 492 | 22 | AAV72559 | Human seripancrin |
| 33 | 657 | 26.8 | 1019 | 22 | AAE06940 | Human enterokinase |
| 34 | 656 | 26.8 | 311 | 21 | AAAB1697 | Mouse serine prote |
| 35 | 655.5 | 26.8 | 416 | 20 | AAV43325 | Mouse hepsin prote |
| 36 | 655.5 | 26.8 | 416 | 20 | AAV96812 | A mouse serine pro |
| 37 | 652 | 26.6 | 432 | 21 | AAV99417 | Human PRO1570 (UNO |
| 38 | 652 | 26.6 | 432 | 22 | AAAB87581 | Human PRO1570 (HO |
| 39 | 652 | 26.6 | 432 | 22 | AAAB66166 | Protein of the inv |
| 40 | 648.5 | 26.5 | 437 | 22 | AAE06931 | Human membrane-ty |
| 41 | 646 | 26.4 | 172 | 22 | AAAM25245 | Human protein sequ |
| 42 | 600.5 | 24.5 | 238 | 21 | AAAB11695 | Mouse serine prote |
| 43 | 583.5 | 23.8 | 317 | 22 | AAAM25633 | Human protein sequ |
| 44 | 576 | 23.5 | 256 | 22 | AAAB98501 | Human hepsin serin |
| 45 | 564.5 | 23.1 | 418 | 17 | AAAB89435 | Trypsin-like enzyme |

ALIGNMENTS

| | | |
|----------|--|--|
| RESULT 1 | AAE06935 | standard; Protein; 453 AA. |
| ID | AAE06935 | |
| XX | AAE06935; | |
| AC | 16-OCT-2001 | (first entry) |
| XX | | |
| XX | | Human membrane-type serine protease (MTSP) 6. |
| DE | | |
| XX | | |
| XX | | Human; transmembrane serine protease; membrane-type serine protease; |
| KW | MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic; | |
| KW | lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| FH | Key | Location/Qualifiers |
| FT | Domain | /label= Transmembrane-domain |
| FT | Domain | /label= LDLR-domain |
| FT | Domain | /note= "LDL receptor domain class a" |
| FT | Domain | /label= SR-domain |
| FT | Domain | /note= "Scavenger receptor Cys-rich domain" |
| FT | Cleavage-site | /note= "Trypsin-like serine protease domain" |
| FT | Domain | 216..443 |
| FT | Domain | 217..443 |
| FT | Domain | /label= Protease-domain |
| FT | Misc-difference | 324 |
| FT | | /note= "Unpaired cysteine" |
| XX | | |
| PN | WO200157194-A2. | |

XX Sequence 453 AA;
SQ
Query Match 99.3%; Score 2430.5; DB 21; Length 453;
Best Local Similarity 99.6%; Pred. No. 6e-187;
Matches 452; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
QY 1 MGENDPAAVAPRPSRSLGIDLDLKITSPVAPDADAQAOLSLPLKFPFIYIGIITLI 60
DB 1 mgenpdpaveapstfslrlgldlklspvpdadavaaqlslplklfpfiyigiall 60
QY 61 IALAIALGIHFDCSGKYRCSSEFKCIELIARCQVSDCKGEDEYRCVVRGONAVLOVF 120
DB 61 lalaialglhfcdsgkyrcsscfkciellarcqvsdckgedeyrcvvrqgnavlvqvf 120
QY 121 TAAASMTKCSDDMKGHYANVACAQLGPPSYSSDNLRVSSLEGQFREFEYSIDHLLPDDK 180
DB 121 taaswktmcsddkghyanavacaglgfpysvssdnlrsvslegqfreefvsidhllpddk 180
QY 181 VTLAHSYVYRREGCAGSHVYTLQCTACGHRGYSRRIVGNSMLSLQMPWQASLQFGYH 240
DB 181 vtlahsyvyrregcagshvylqctacghrgrysrrivgnsmlslqmpwqaslgfgyyh 240
QY 241 LCGSVITPLMITTAACHVYDLYLPKSWITQVGLVSLDNPAPSHLYEKIVYHSKTKPKR 300
DB 241 lcgsvitpmittaachvydlylpkswitqvglvslidnpapshlyekivyhsktkpkkr 300
QY 301 LGNDILMLKIAGPLTFNEMIQPVCLPNSSENPDPGKVCWTSGMGATDAGDASPVLNHA 360
DB 301 lgndilmlkiagpltfnemiqpvcclpnssenpdpgkvcwtsgmgatdagdaspvlnha 360
QY 361 AVPLISNKNICNHRDYGVIITSPSMLCAGYLTGVDSCQSGSGRPVCOEAEKIKVIGATS 420
DB 361 avplisnknicnhrdygviitspsmcagyltgvdscqsgsgprvcoeaeekikvigats 420
QY 421 FGIGCAEVNKPQYTRVTSFLDWIHEQMERDLKT 454
DB 421 fgigcaevnkpqytrvtstfldwiheqmerdlkt 453
RESULT 3
ID AAY41694 standard; Protein; 452 AA.
XX AAY41694;
AC
XX
DT 07-DEC-1999 (first entry)
XX
XX Human PRO382 protein sequence.
DE
XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
secreted protein; transmembrane protein.
XX
XX Homo sapiens.
XX
PN WO9946281-A2.
XX
PD 16-SEP-1999.
XX
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.

PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 13-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.

PR 28-MAY-1998; 98US-0087098.
 PR 28-MAY-1998; 98US-0087106.
 PR 28-MAY-1998; 98US-0087208.
 PR 30-JUL-1998; 98US-0094651.
 PR 11-SEP-1998; 98US-0100038.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
 XX
 DR WPI: 1999-551358/46.
 DR N-PSDB; AA233949.
 XX
 PT New secreted and transmembrane polypeptides and their polynucleotides,
 PT useful for treating blood coagulation disorders, cancers and cellular
 PT adhesion disorders -
 XX
 PS Claim 12; Fig 28; 530pp; English.
 XX
 CC The present invention describes secreted and transmembrane polypeptides
 CC and their polynucleotides. The nucleotide sequences are useful as
 CC sources of probes, primers, for chromosome mapping, and for generation
 CC of antisense sequences. They can also be used to create transgenic
 CC animals. The proteins can be used to treat a variety of diseases and
 CC disorders, depending on their function. Diseases that may be treated
 CC include blood coagulation disorders, cancers and cellular adhesion
 CC disorders. They may also be used to raise antibodies. AA233891 to
 CC AA234338, and AA41685 to AA41774 represent polynucleotide and
 CC polypeptide sequence given in the exemplification of the present
 CC invention.
 XX
 XX Sequence 452 AA;
 SQ

Query Match 98.8%; Score 2418; DB 20; Length 452;
 Best Local Similarity 99.6%; Pred. No. 6e-186; 0; Indels 2; Gaps 2;
 Matches 452; Conservative 0; Mismatches 0;

QY 1 MGENDPPAVAPSPFSRSLFGIDDLKISPVADDAVAQAQLSLPLKFPPIVIGITALI 60
 DB 1 mgendppaveapfsrslfglddlkispvaddavaqaqlslplkfpplvlgitali 59
 QY 61 LALATLGLGHFDCSGKRYRCRSSFKCIELARCDGVSCKDGEDEYRCVYRGONAVLOY 120
 DB 60 lalatlglghfcdsgkryrcrssfkciellarcdgvsckdgedeyrcvrygqnavloy 119
 QY 121 TAAAWKTCSDDMKGYHANYACAOLGFPYSSDNLRVSSLEGOFEFYSIDHLLPDDK 180
 DB 120 taaswtmcscddmkgyhanyacacaglgfpyssdnlrvssleqgfreesidhllpddk 179
 QY 181 VTALHSHSVYREGCASHVVTLOCTACGHRGYSRRIYGNMSILSQPMQASLQFOGYH 240
 DB 180 vtalhshsvyregcasghvvtlqctacghrgrysrriygnmsilseqpmqaslgfgyh 239
 QY 241 LCGGSVTPETPMTTAAACVVDLTLPKSWTQVGLVSLDMPAPSHLYEKYVHSHKPKR 300
 DB 240 lcggsvtpetpmtttaacvvdltlpskwtqvglvslldmpapshlyekyvshkpkrr 299
 QY 301 LGGNDIALMKIACPLTENEMIQPCLPNSSENFDPGKVCMTSGMGATEGAGDASPVLNHA 360
 DB 300 lggndialmkiaapltenemiqpclpnssefnfdpgkvcwtsmgatgedg_gdaspvlnha 358
 QY 361 AVPLISNKCINHRDVGIGIISPSMLCAGYLTGVVSCQSGSGGLVCOEERLMLKVAGTS 420
 DB 359 avplisnkcicnhrdvygigilspsmcagyltgvdscgsgsgplvcgerllwklvgats 418
 QY 421 FGIGCAEVNKRKYVTRVTSFLDWHHEQMERDLYT 454
 DB 419 fgigcaevnkrpyvtrvtsfldwhheqmerdlyt 452

RESULT 4
 AAB32246

ID AAB32246 standard; Protein: 454 AA.
 XX
 AC AAB32246;
 XX
 DT 11-JAN-2001 (first entry)
 XX
 DE Tumour associated differentially-expressed gene 12 protein sequence.
 XX
 KW Transmembrane serine protease; TADG-12; chromosome 17; vaccination;
 KW tumour associated differentially-expressed gene 12; cytotaxtic; human;
 KW malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
 XX
 OS Homo sapiens.
 XX
 PN WO200052044-A1.
 XX
 PD 08-SEP-2000.
 XX
 PF 02-MAR-2000; 2000WO-US05612.
 XX
 PR 03-MAR-1999; 99US-0261416.
 XX
 PA (UYAR-) UNIV ARKANSAS.
 XX
 PI O'Brien TJ, Underwood LJ;
 XX
 DR WPI: 2000-533263/48.
 DR N-PSDB; AAA93842.
 XX
 PT DNA fragment encoding tumor associated differentially-expressed gene 12
 PT protein used for diagnosing and treating malignant hyperplasia and
 PT cancers including ovarian cancer -
 XX
 PS Claim 3; Figure 4; 118pp; English.

CC This invention relates to a novel transmembrane serine protease called
 CC tumour associated differentially-expressed gene 12 (TADG-12). TADG-12 is
 CC located on chromosome 17. Sequences AA93842-A93845 and AAB32246-B32249
 CC represent human TADG-12 cDNA and their corresponding protein sequences.
 CC A splice variant of TADG-12 (TADG-12V) leads to a truncated protein
 CC product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12
 CC exhibits cytosolic activity, and can be used in vaccines and in gene
 CC therapy. TADG-12 nucleotide and protein sequences are used in the
 CC diagnosis of malignant hyperplasia and cancers of the ovary, breast,
 CC lung, colon, prostate and other cancers where TADG-12 is overexpressed.
 CC TADG-12 is particularly used as tumour marker for early disease
 CC diagnosis. TADG12 proteins or fragments can be used to vaccinate an
 CC individual with cancer, suspected of having a cancer or at risk of
 CC getting cancer. Sequences AA93846-A93853 represent PCR primers used for
 CC amplifying the TADG-12 cDNA sequence, and in the quantitative analysis of
 CC TADG-12 mRNA. AAB32250 represents a peptide fragment of TADG-12, used to
 CC create anti-TADG-12 antibodies. Sequences AAB32251-B32269 represent
 CC TADG-12 peptides which target HLA, and may be used in a vaccine or for
 CC immune stimulation.
 XX
 XX Sequence 454 AA;
 SQ

Query Match 97.1%; Score 2375; DB 21; Length 454;
 Best Local Similarity 97.6%; Pred. No. 1.7e-182;
 Matches 444; Conservative 2; Mismatches 7; Indels 2; Gaps 2;

QY 1 MGENDPPAVAPSPFSRSLFGIDDLKISPVADDAVAQAQLSLPLKFPPIVIGITALI 59
 DB 1 mgendppaveapfsrslfglddlkispvaddavaqaqlslplkfpplvlgitali 60
 QY 60 ILALATLGLGHFDCSGKRYRCRSSFKCIELARCDGVSCKDGEDEYRCVYRGONAVLOY 119
 DB 61 ilalatlglghfcdsgkryrcrssfkciellarcdgvsckdgedeyrcvrygqnavloy 120
 QY 120 FTAASWKTMCSDMKGYHANYACAOLGFPYSSDNLRVSSLEGOFEFYSIDHLLPDD 179
 DB 121 ftaaswtmcscddmkgyhanyacacaglgfpyssdnlrvssleqgfreesidhllpdd 180

QY 180 KVTALHSHVYVREGASGHVVTLOCTACGHRGYSRIVGNNLSLQPMQASLOFGY 239
 |||||||
 Db 181 kvcalhshsvyregcasghvvtlqctacghrgrysriviagmmsllsqwpgasldfgy 240
 QY 240 HLCGGSVITPMTTAAHCYVDLTPKSWTIQVGLVSLDNPAPSHLVEKIVYHSKPKR 239
 |||||||
 Db 241 hlcggsvitpmtttaahcyvdltpkswtiqvglvslldnpapshlvekiyhsykp 300
 QY 300 RLGNDAIIMKLAGPLTFENEMIQPVCLPENSENFPDGKVCMTSGMGATBDGAGASPLNH 359
 |||||||
 Db 301 rlgnndaimklagpltfenemiqpvcipnseemfpdgkvcwtsygatdeg-qdaspylm 359
 QY 360 AAVPLISNKNHNDYVGGIISPSMLCAGYLTGVDSCQDSCGPIVQERRIMKLVGAT 419
 |||||||
 Db 360 aavplisnknhndryvggilspsmlcagyltgvdscqdsqgplvcqerrilwklvgat 419
 QY 420 SFGICAEVNNKPGVYTRVTSFLDWIHEQMERDLKT 454
 |||||||
 Db 420 sfigicaevnkpvytrvtsfldwlheqmerdlkt 454

RESULT 5
 AAY72093 standard; Protein; 327 AA.
 XX
 AC AAY72093;
 XX
 DT 28-MAR-2001 (first entry)
 XX
 DE Human serine protease #4 encoded by clone HCHAK72.
 XX
 KW Human: serine protease; osteopathic; immunosuppressive; anti-allergic;
 KW anti-inflammatory; cytoskeletal; cardiac; neuroprotective; nootropic;
 KW neuroleptic; vulnereary; ophthalmological; antibacterial; antiviral;
 KW antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma;
 KW treatment; bone formation disorder; osteoporosis; arthritis; cancer;
 KW connective tissue disorder; autoimmune disorder; wound healing; asthma;
 KW systemic lupus erythematosus; male reproductive system disorder;
 KW testicular cancer; digestion and food absorption disorder; arrhythmia;
 KW Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy;
 KW behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;
 KW cardiovascular disorder; ocular disorder; drug screening.
 KW
 OS Homo sapiens.
 XX
 PN WO200068247-A2.
 XX
 PD 16-NOV-2000.
 XX
 PF 05-MAY-2000; 2000MO-US12207.
 XX
 PR 07-MAY-1999; 99US-0133239.
 PR 20-MAY-1999; 99US-0135163.
 PR 03-AUG-1999; 99US-0147005.
 PR 09-SEP-1999; 99US-0152935.
 PR 01-NOV-1999; 99US-0162979.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Shi Y, Young PE, Ni J;
 XX
 DR WPI; 2000-67979/66.
 DR N-PSDB; AAD02323.
 XX
 PT New nucleic acid molecules encoding human serine protease polypeptides,
 PT useful for diagnosis, prevention and/or treatment of disorders e.g.
 PT osteoporosis, lupus erythematosus and Alzheimer's -
 XX
 PS Claim 12; Page 280-281; 289pp; English.
 XX
 CC The present sequence is human serine protease #4 from clone
 CC HCHAK72 (ATCC Deposit No: PTA27).

CC The invention relates to human serine proteases and their cDNA clones.
 CC It is used in methods for the diagnosis, prevention and treatment of
 CC various disorders related to serine protease such as bone formation
 CC disorders (osteoporosis), connective tissue disorders (arthritis),
 CC autoimmune disorders (systemic lupus erythematosus), wound healing, male
 CC reproductive system disorders (testicular cancer), digestion and food
 CC absorption disorders (Crohn's disease), neurodegenerative diseases
 CC (Alzheimer's disease), behavioral disorders (Tourette's syndrome),
 CC proliferative and cancerous conditions (acute myelogenous leukaemia),
 CC allergic reactions (asthma), cardiovascular disorders (arrhythmia),
 CC ocular disorders (glaucoma) and infectious diseases caused by bacteria,
 CC viruses, fungi or parasites. It is also useful for screening therapeutic
 CC compounds. Serine proteases are used as immunological probes or
 CC polymorphic markers for the identification of chromosomes, cells and
 CC tissues in biological samples, identification of male contraceptive
 CC agents, delivery of compositions to targeted cells expressing a
 CC receptor for serine protease, hybridisation probes and molecular weight
 CC markers. Serine protease nucleic acids are also useful in gene therapy.
 CC Note: The present sequence shown in page 280-281 of sequence listing has
 CC been assigned SEQ ID NO: 14. But the sequence, human serine protease
 CC epidermal growth factor (EGF)-like domain (AAY72114) shown in
 CC page 12 is also referred as SEQ ID NO: 14.
 CC
 SQ XX Sequence 327 AA;

Query Match 73.1%; Score 1788; DB 21; Length 327;
 Best Local Similarity 100.0%; Pred. NO. 1.4e-135;
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 MCSDDMKGHYANVACAQLGPPSYVSDNLRVSSLGQFREFVYSIDLPLDDKVTALHHS 187
 |||||||
 Db 1 mcsddmkghyanvacagqlgfpsyvssdnlrvslegfrefeysidhllpddkvtalhs 60
 QY 188 VYVREGCASHSVVTLOCTACGHRGYSRIVGNNLSLQPMQASLOFGYHLCGGSVI 247
 |||||||
 Db 61 vyvregcasghvvtlqctacghrgrysriviagmmsllsqwpgasldfgyhlcggsvl 120
 QY 248 TPLMTTAAHCYVDLTPKSWTIQVGLVSLDNPAPSHLVEKIVYHSKPKRLGNDAI 307
 |||||||
 Db 121 tplmtttaahcyvdltpkswtiqvglvslldnpapshlvekiyhsykpkrilgnndal 180
 QY 308 MKLAGPLTFENEMIQPVCLPENSENFPDGKVCMTSGMGATBDGAGASPLVNHAAVPLISN 367
 |||||||
 Db 181 mklagpltfenemiqpvcipnseemfpdgkvcwtsygatdegqdaspylnhaavplism 240
 QY 368 KICNHRDYYGGIISPSMLCAGYLTGVDSCQDSCGPIVQERRIMKLVGATSGFGCAE 427
 |||||||
 Db 241 kicnhrdyvggilspsmlcagyltgvdscqdsqgplvcqerrilwklvgatstfgycae 300
 QY 428 VNKPGVYTRVTSFLDWIHEQMERDLKT 454
 |||||||
 Db 301 vnkpgvytrvtsfldwlheqmerdlkt 327

RESULT 6
 AAB32247 standard; Protein; 294 AA.
 XX
 AC AAB32247;
 XX
 DT 11-JAN-2001 (first entry)
 XX
 DE Tumour associated differentially-expressed gene 12 variant 3 protein.
 XX
 KW Transmembrane serine protease; TABD-12; chromosome 17; vaccination;
 KW tumour associated differentially-expressed gene 12; cytoskeletal; human;
 KW malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
 XX
 OS Homo sapiens.
 XX
 PN WO200052044-A1.
 XX

QY 330 ENPDKVCWTSCMGATIEDGADASPYLNHAAVPLISNKNICNHRDVGIIISPSMLCAGY 389
124 enfpdgkvwtscgwtsgatedgagdaspylnhaavplisnknichrdvgygiiispsmlcagy 183
Db 390 LTGVDSGCGSDSGPLVCQERRLMKLVGATSFEGICAEVKNPGVYTRVTSFLDWMIEOME 449
184 ltgvdscgsgdsgplvcqerlmlklygatsfigicaevnkpvytrvtsfldwmieome 243
QY 450 RDLKT 454
Db 244 rdlkt 248

RESULT 8
AAG73966
ID AAG73966 standard; Protein; 193 AA.
AC AAG73966;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:4730.
XX
KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
PD 05-APR-2001.
PF 28-SEP-2000; 2000WO-US26524.
XX
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR N-PSDB; AAH33397.
XX
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 6530-6531; 9803pp; English.
XX
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing P.
CC Inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated PS,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX
SQ Sequence 193 AA;

Query Match 42.3%; Score 1034; DB 22; Length 193;
Best Local Similarity 100.0%; Pred. No. 2.2e-75;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 PKSMTOVGLVSLDNPAPBHLVEKIVYHASKYPRKRGNDIALMKLAGPLTFEMIOVC 324
4 pkswtlqvglvslldnpapbhlveklyvhaskyprkrgndialmklagpltfemioqpc 63
Db 325 LPNSENFPDQKVCWTSCMGATIEDGADASPYLNHAAVPLISNKNICNHRDVGIIISPSM 384
64 lpnseentfpgdkvwtscgwtsgatedgagdaspylnhaavplisnknichrdvgygiiispsm 123
QY 385 LCAGYLTGVDSGCGSDSGPLVCQERRLMKLVGATSFEGICAEVKNPGVYTRVTSFLDWM1 444
124 lcaagyltgvdsdgsgdsgplvcqerlmlklygatsfigicaevnkpvytrvtsfldw1 183
Db 445 HEOMERDLKT 454
184 heqmerdlkt 193

RESULT 9
AA92050
ID AA92050 standard; Protein; 492 AA.
AC AA92050;
XX
DT 01-AUG-2000 (first entry)
XX
DE HRPc6/7 polypeptide from androgen-inducible gene clone.
XX
KW Androgen inducible; testosterone; prostate cancer; cytostatic;
KM TMPRS2; diagnosis.
XX
OS Homo sapiens.
XX
PN WO200018961-A2.
PD 06-APR-2000.
PF 30-SEP-1999; 99WO-US22535.
XX
XX
PR 30-SEP-1998; 98US-0163759.
PR 30-SEP-1998; 98US-0164159.
XX
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Macbeth KJ, Shyjan AW;
XX
DR WPI; 2000-293182/25.
DR N-PSDB; AAA08803.
XX
XX
PT Novel methods for identifying compounds for treating prostate cancer
PT comprising measuring the level of expression or activity of 1 or more
XX
XX
PS Claim 2; Fig 3; 108pp; English.
XX
XX
CC This protein is encoded by a gene which is androgen (e.g. testosterone)
CC inducible in androgen-dependent prostate cancer cells (e.g. LNCap cells)
CC and constitutively expressed in androgen-independent prostate cancer
CC cells (e.g. LNCap cells). Agents which decrease the expression or
CC activity of these clones may slow or arrest the growth of prostate cancer
CC cells or may kill them. HRPc6/7 can be obtained from the sequence of
CC the known gene for TMPRS2. A compound useful for treating prostate
CC cancer can be identified in a novel method comprising measuring the
CC expression level, or activity, of HRPc2, 3, 6/7, 8, 9, 10, 13, 14, 15,
CC 19, or peripheral-type benzodiazepine receptor (PBR) in a cell, in the
CC presence and absence of a test compound. The sequences may also be used
CC in diagnosis of prostate cancer and to determine efficacy of treatment
CC for prostate cancer.
XX
XX
SQ Sequence 492 AA;

PA (MYRIAD GENETICS INC.
 PI Wong AKC, Tavtigian SV, Teng DHF;
 XX WPI; 2000-170914/15.
 DR N-PSDB; AA287786.
 XX
 PT Novel tumor suppressor TMRSS2 used for the diagnosis and prognosis of
 PT human cancer -
 PS
 PS Claim 55; Page 77-79; 89pp; English.
 XX
 XX The invention provides a new tumor suppressor gene, designated TMRSS2.
 CC The TMRSS2 polynucleotides and polypeptides can be used in methods for
 CC diagnosing and prognosing predisposition to cancer in humans. The
 CC polypeptides may also be used in assays to screen for compounds with
 CC anti-cancer or therapeutic properties. The polypeptides are also useful
 CC for rational drug design. The TMRSS2 polynucleotides and polypeptides
 CC may be used for gene therapy and protein therapy. The present sequence
 CC represents the TMRSS2 polypeptide.
 XX
 XX Sequence 492 AA:
 50

| | | | | |
|---------------------------|-------|--------------------|------------|-------------|
| Query Match | 35.8% | Score 875.5; | DB 21; | Length 492; |
| Best Local Similarity | 45.9% | Pred. No. 3.5e-62; | | |
| Matches 189; Conservative | 53; | Mismatches 149; | Indels 21; | Gaps 11 |

```

OY 52 IYIGIIIALIIALAIIGIHF---DCSGK-YRORSSPKCIELARCOGVSDCKDGEDEYRC 107
Db 89 lltlgtfitygaalaagllwfkngskscnsngiecdsgtclnpsncwcvghshcpgdeicntc 148
OY 108 VAVGQOMANVLOYFTA--ASWTKTCSDDMKGHYANVACAOIGFP-SYVSSDNIIRVSSLEQO 164
Db 149 vrlgpnffllgysssqgrkswhpvcqddmwenygractcmg/knfnfysqglivddsgts 208
OY 155 FREEVVSIDHLPDKVATALHHSVYVRBSCASGHVVTLQCTACGHRGYS--SRIYGMN 222
Db 209 fmkltntsgnv---dluykklhns---dacsckavavslrlcagvlnnsrgsrliygves 261
OY 223 SLLSQWPMOASLQFQGYHLCGSGVSTPLIITIAAHCV-YDLVLPKSKWTLQVGLV--SLLD 279
Db 262 alpgawpavqslhvnqvnvncgsglltlpewlyaaahcvekrplnmpwhtefagllrgstlmf 321
OY 280 NPAPSHLEKTIYVHSHKRYKPKRLGNIDIALMLKLAGPLTFNEMIOVCLPNSSEENFPDDKVCW 339
Db 322 yga-glyevexvlshpnyscktkmndatlmkldqrltfndlvrcvclpmpngmmldqpsqelc 380
OY 340 TSGCATEDGACDASPVLNHAAVPLISNKTICNHRDVGYSISPSMLCAGYLRTGVDSGCG 399
Db 381 tsywgatgee-kgtsevlinaakvlllletdrcnsryyudnlltpramicagfllgnavdscg 439
OY 400 DSGGGLVQOERLMLKLTGATSFGICCAEVRNKGVYTRVTSFLDMIHQEQERD 451
Db 440 dsngglvsknmliwlldtswgscakaxayrgvynvnmvftcdwvlgqrmd 491

```

RESULT 12
AAV44406
ID AAV44406 standard; Protein; 492 AA.
XX
XX
AC AAV44406;
XX
XX
DT 22-MAR-2000 (first entry)
XX
XX
DE Human 20P1F12-GTC2 protein.
XX
XX
KW 20P1F12; TMPRSS2; androgen; serine protease; 20P1F12-GTC1; cancer;
transmembrane protein; colon; prostate; prostate tumour,
XX
XX
OS Homo sapiens.
XX
PN W09962942-A2.

| | |
|----|--|
| XX | 09-DEC-1999. |
| PD | |
| XX | |
| PF | 01-JUN-1999; 99WO-US12253. |
| XX | |
| PR | 01-JUN-1998; 98US-0087598. |
| PR | 29-JUN-1998; 98US-0091474. |
| PR | 14-APR-1999; 99US-0129521. |
| XX | |
| PA | (UROG-) UROGENESYS INC. |
| PA | (AFAR/) AFAR D E. |
| PA | (HUBE/) HUBERT R S. |
| PA | (LEON/) LEONG K. |
| PA | (RAIT/) RAITANO A B. |
| PA | (SAFE/) SAFFRAN D C. |
| XX | |
| PI | Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC; |
| XX | |
| DR | WPI; 2000-116363/10. |
| DR | N-PSDB; AAZ29636. |
| XX | |
| PT | Novel cell surface antigen useful to treat colon and prostate cancer - |
| PS | Claim 1; Fig 1; 58pp; English. |

| | | | | | | | |
|-----------------------|--------|--------------|--------|------------|-----|--------|-----|
| Query Match | 35.8% | Score | 875.5 | DB | 21 | Length | 492 |
| Best Local Similarity | 45.98% | Pred. No. | 3,5662 | | | | |
| Matches | 189 | Conservative | 53 | Mismatches | 149 | Indels | 21 |
| | | | | | | Gaps | 11 |

```

OY 52 IVIGIALIALALGIGHNF---PCSG-RYCRSSFKCIETILACDDGDCDKEDEYRC 107
Db 89 ltlglfivgaalaagllwfkmgskscmsnglcedssgyltlnpsmcodyshpbgdeanrc 148
OY 108 VRVGQNAVLYOFMA--ASMKTCSDDMKGYHANYACAQLGP--SYVSSDNLRVYSLSEQ 164
Db 149 vrllypnfllygysqgrkshwpvcqddnmenygraacrdmgymknfysagflvddsgts 208
OY 165 FREEVNSIDLHLPPDDKVTALHSHSYVRGGCSAGHVITLOCTACGHRKRG--SKLYGGM 222
Db 209 fmkIntesagv--dlykklYhs---dacsckavslrlciacgvnljnsrsgrflvges 261
OY 223 SLLSQWPAQASLQFGYHLLGSGSVTTPRLMITTAHCV-YDYLKRSWTLQYGLV--SLLD 279
Db 262 alpgawqvgsllhvqnvnvccgslltpewlvtlaacvcekrllnmpwhvtaefatllrgsfmf 321
OY 280 NPAASHLEKELIVYHSKYPKPKRLGNDIALMKLAGPLTFENEMIQPCLPSPSEENFPDGKCW 339
Db 322 yga-gygcexkrlshpnyscktkkndialmkqkrlftfdlvykpcclprpmmllpcedqclw 380
OY 340 TSGWCATEDGAGDASPYVLNHAAPRLISNKTGNHRDYYGCIISPSMLCAGYLITGCVDSQCG 399
Db 381 lsgwgatee-kgktssevlnaackvlllietqrcnsryyudllitpemi cegflgvnvcsg 439
OY 400 DSGGRVLVQDERKLMKLVGATSFGLGCAEVNKPYYTRKTSFLMDIINHEMEND 451
Db 440 dsagrpvltskmlwlllgdtlswsgcaakayppryugvnmvltfdwlyqrmad 491

```

RESULT 13
AA01317
ID AA01317 standard; Protein; 393 AA.
XX
AC AA01317;
XX
DT 04-OCT-2001 (first entry)
XX
DE P1000C partial amino acid sequence 100-492.
XX
KM Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytosolic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kales MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAM;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Claim 2; Page 540-541; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used to treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (II) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AA01115 to
CC AA01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 393 AA.

Query Match 35.5%; Score 868.5; DB 22; Length 393;
Best Local Similarity 46.6%; Pred. No. 9.7e-62;
Matches 187; Conservative 50; Mismatches 143; Indels 21; Gaps 11;
OY 63 LAIGLGHF---DCSGK-YRCRSSFKCIELIARCDGVSDCKDEDEYRCYRGGAQNAVQ 118
DB 1 laagllwktmgsksgscngicldspnwcgvshcpqgdenrcvrlgysnfllq 60
OY 119 VFPA--ASWKTMCSDMKHGVANVACAOLGFP-SYVSSDNLARSSLEQGFREFEVIDH 175
DB 61 vvsqqrkshpvcqddmnenygraactdmgykntfysgqivddsgstsfmkntsaagnv 120
OY 176 LPDDKVTALHHSVYVEGCAGSHVYTLQCTACGHRRGYS--SRIVCGNNSILSQMPQAS 233
DB 121 ---dlykklkylhs-----daccskavslrclacgvlnnsrgrstiyggesalpgapwpqys 173
OY 234 LQFGYHLCGGSVITPLWITTAHCY-YDLYLPRKSWTIOGVLV--SLLDNPARSHLVERI 290

DB 174 lhvgnhvcgssiltpewivlaahcvekpdlmgwhwtafagllrgsfmfyga-gyqvekv 232
OY 291 VYHSKXKPKRLGNDILMKIAGLTFNEMIQPCLCNSENFDPDGKVCWCMGSGCATIEDGA 350
DB 233 lshpnyskstkndialmkkpkpltdlvkpyclpnpqmmllqpeqlcwlsgvatlee-k 291
OY 351 GDASPVLMNHAAPVPLNSKICNHRDVGILISPSMLCAGYLTGVDSCQDGSGLPVCQER 410
DB 292 gkxsevlinaakvllietqrnsryyddnlitpmicagflqgnvdscgdsqgplvtskn 351
OY 411 RLMKLVGATSEFGICAEVKNKPGVYTRVTSFLDIHROMERD 451
DB 352 nlwvlllgdtswsgscakayrpygygvnmvftdwlryqmrad 392

RESULT 14
AAB36901
ID AAB36901 standard; Protein; 492 AA.
XX
AC AAB36901;
XX
DT 26-FEB-2001 (first entry)
XX
DE Human TMPRSS2 protein.
XX
KM Prostate specific androgen regulated protein; ARSRL; TMPRSS2;
KW PART-1; neoplastic.
XX
OS Homo sapiens.
XX
PN WO200065067-A2.
PD 02-NOV-2000.
XX
PF 21-APR-2000; 2000WO-US10920.
XX
PR 23-APR-1999; 99US-0130778.
PR 30-AUG-1999; 99US-0151585.
PR 30-DEC-1999; 99US-0174003.
PR 24-JAN-2000; 2000US-0177751.
XX
PA (UNITV) UNITV WASHINGTON.
XX
PI Nelson PS, Hood L, Lin B;
XX
DR WPI; 2000-679676/66.
XX
DR N-PSDB; AAC83325.
XX
PT Polynucleotide encoding prostate specific androgen regulated
PT polypeptides and inhibitor of the peptides useful for treating or
PT reducing the progression of prostate neoplastic condition in an
PT individual -
XX
PS Claim 63; Page 86-88; 121pp; English.
XX

Query Match 35.3%; Score 864.5; DB 21; Length 492;
Best Local Similarity 45.4%; Pred. No. 2.6e-61;
Matches 186; Conservative 55; Mismatches 146; Indels 21; Gaps 11;
OY 52 IVIGIALLALAIIGLGHF---DCSGK-YRCRSSFKCIELIARCDGVSDCKDEDEYRC 107

Thu Apr 11 09:27:12 2002

us-09-846-512-12.rag

Page 12

DB 149 VRLGPNFLLQVXSORKSWHPVQDDMNMENYGRACRDMGKKNFESSQGIIVDSGSTS 208
 QY 165 FREEFIDHLLPDDKVTALHSHVYREGCASHVVTLOCTAGCHRGYS--SRIYGNM 222
 DB 209 FKLNTSAGNV--DIYKKLYHS---DACSAAVSLRCLAGVNLNSRSRIVGES 261
 QY 223 SLISQMPQASLOFGYHLGCGSVITPLMTITTAHCV-YDLYLPKSWTIOVGLV--SLLD 279
 DB 262 ALPGAMPQVSLHGVNVHVGCGSIITPEWITVAHCEKPLNPMWMTAFAGILRSQFMF 321
 QY 280 NPAPSHLVKVIYHSKYPKRLGNDIALMKLACPLTFNEMIQVCLPNSSENFDDGKVCW 339
 DB 322 YGA-GQVKEVYSHPNYDSSTKNNIDALMKLQKPLTFENDLVKPCLPNPGMLQPDQLCW 380
 QY 340 TSGMGATEGAGDASPVLNHAAPVLISNKNCRDYGGIISPSMLCAGYLTGVDSCOG 399
 DB 381 ISGMGATEE-KGKTEVYNAAKAVLLIETORCNSRYVDNLITPAMICAGFLQGNVSCOG 439
 QY 400 DSGGPLVCOERRLMKLVGATSFEGICAEVKNKPGVYTRVTSFLDMIHOMERD 451
 DB 440 DSGGPLVTSKNNIMWMLIGDTSMGSCAKAYRPGVGNVWFTDMYIROMRAD 491

RESULT 2

Q9DGR2 PRELIMINARY: PRT; 767 AA.
 AC Q9DGR2;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE EMBRYONIC SERINE PROTEASE-2.
 GN XESP-2.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20363741; PubMed-10903452;
 RA Yamada K., Takabatake T., Takeshima K.;
 RT "Isolation and characterization of three novel serine protease genes
 from Xenopus laevis."
 RL Gene 252:209-216(2000).
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
 (LDLRA) DOMAIN.
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 CC EMBL; AB038497; BAB08217.1; -;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LDL_receptl_A.
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam; PF00057; ldl_receptl_a; 6.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00192; LDLa; 8.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS00068; LDLRA_2; 2.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN.1.
 KW Glycoprotein; Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE96E CRC64;

Query Match 33.7%; Score 825.5; DB 13; Length 767;
 Best Local Similarity 45.0%; Pred. No. 5,7e-69;
 Matches 175; Conservative 46; Mismatches 131; Indels 37; Gaps 10;
 QY 78 RCRSFKCIELIARCDVSDCKDGEDERYCVRVGNNAVLYQV--TAASMKTCSDMKG 135
 DB 394 RCGSSVSVLSSQWCDGVSDCPYGEDEMGSVSLYIPADFOLQVYSTSVSAMLVPCSDYWND 453

QY 136 HYANVACAQLGF--PSYVSSDNL-----VSLGQREEFVSDHLLPDDKYTALH 185
 DB 454 DFRFRCQDFGIGSSYNNRDTLMSYPANPGYFKLYSGWRKSF-----YTSVQ 502
 QY 186 HSYVYREGCASHVVTLOCTACG-HRRGYSSRIYGVGNMSLLSQMPQASLOFGYHLGCG 244
 DB 503 YSSY-----CVSGNVSLHCSICVSNNSLVSRIYVGTFNALGMWPMQVNLQITGVLCG 558
 QY 245 SVITPLMTITTAHCVVDLYPKS-WTIQVGLVLLDNF-----APSHLVKVIYHSKYPK 299
 DB 559 SLISPMWITVAHCVYGSVSSASGMRVFAG--TLTKPSYNSAAVEERIVHPGKSY 615
 QY 300 RLGNDAIMKLACPLTFNEMIQVCLPNSSENFDDGKVCWTSGMGATEGAGDASPVLN 359
 DB 616 TYNDIALMKLRDEITFGYITQVPCLPNSGMFEAGTTTWISSMGSTYEG-GSVITYLOY 674
 QY 360 AAVPLISNKNCRDYGGIISPSMLCAGYLTGVDSCOGDGGPLVCOERRLMKLVGAT 419
 DB 675 AAIPLIDSNVNCOSYVNGQITSSMICAGYLSGVDTCGDSGGLPNKRNKGTWMLVGD 734
 QY 420 SFGICAEVKNKPGVYTRVTSFLDMIHOM 448
 DB 735 SWDGCARANKPGVGNVTFLEWISQM 763

RESULT 3

Q9BYE2 PRELIMINARY: PRT; 581 AA.
 AC Q9BYE2;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE MEMBRANE-TYPE MOSAIC SERINE PROTEASE.
 GN MSPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RX MEDLINE-21167393; PubMed-11267681;
 RA Kim D.R., Sharmir S., Inoue M., Kido H.;
 RT "Cloning and expression of novel mosaic serine proteases with and
 without a transmembrane domain from human lung."
 RL Biochim. Biophys. Acta 1518:204-209(2001).
 DR EMBL; AB048796; BAB39741.1; -;
 KW Protease.
 SQ SEQUENCE 581 AA; 62689 MW; 4DABE24D7D5BA44A CRC64;

Query Match 29.8%; Score 728; DB 4; Length 581;
 Best Local Similarity 37.9%; Pred. No. 6e-60;
 Matches 157; Conservative 66; Mismatches 163; Indels 28; Gaps 9;

QY 44 LPLKFRPIIVIGIALILALIGLGHPC-----SKYRCRSFKCIELIARCDVSDCK 99
 DB 161 LPL-----IGVLLILALVSLILFLQFGHTRIRKEQRESCPKAAVRCDDGVDDK 213
 QY 100 DGEDEYRCVYRGVGNNAVLFVTAAS--WRTMCSDDMKGYANVACQQLGFPVSSDNL 157
 DB 214 LKSDELGCVRFMDKSLILTIYSSSHQWLPICSSNNNDSTSEKTCQLQGFESNHTTEVA 273
 QY 158 VSLGQFREEFVSDHLLPDDKYTALHSHVYREGCASHVVTLOCTAGHRRGYSSRI 217
 DB 274 -----HRDFANSFILRYN--STIQESLH-RSHCPQORYISIQCSHCQ-LRAMTGR 321
 QY 218 VGSNMSLLSQMPQASLOFGYHLGCGSVITPLMTITTAHCVYDL--YLPKSWTIOVGL 275
 DB 322 VGGALNDSKRWQVQVSLHFGTTHIGGLIDQWVLTAAHCFEYVREKYLEGKYYAGTS 381
 QY 276 SLIDNPAPSHLVKVIYHSKYPKRLGNDIALMKLACPLTFNEMIQVCLPNSSENFDDG 335

DB 382 NH0LPEASIAE-IIINSNTDEEDDYDIALMLRLSKPLTSLAHIPACLPMMHGTFFSLN 440
QY 336 KYCMTSGCATDGDGADSPVLNHAAPLISNKICNHRVYGGIISPMILCAGYLTGVD 395
DB 441 ETCWITGFGKTRTDKTSPLREVONLIDFKCNDYLVYSSYLLPRMKAGDLHGHD 500
QY 396 SCQSGSGGLVCOERLMLKVGATSGFICAEVKNKPGVYTRVTSFLDWIHQOME 449
DB 501 SCQSGSGGLVCOERLMLKVGATSGFICAEVKNKPGVYTRVTSFLDWIHQOME 554
RESULT 4
Q99BE1 PRELIMINARY: PRT: 537 AA.
AC Q99BE1: 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE MOSAIC SERINE PROTEASE.
GN MSPS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmah S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
RT without a transmembrane domain from human lung."
RL Biochim. Biophys. Acta 1518:204-209(2001).
DR EMBL: AB048797; BAB39742.1;
KW Protease.
SQ SEQUENCE 537 AA: 58102 MW: A39F4E8B16DAECF CRC64;
Query Match 29.1%; Score 712.5; DB 4; Length 537;
Best Local Similarity 38.9%; Pred. No. 1.5e-58;
Matches 145; Conservative 64; Mismatches 147; Indels 17; Gaps 7;
QY 85 CIELIACDGVSDCKDGEDRRCVVGONAVLOVFTAS--WKTWCSDMKGHVAVAC 142
DB 174 CRKHAHRCGVVDCKLSPDELGVPRDMKSLKITYSGSHQWMLPICSNMNDVSEKTC 233
QY 143 AQLGFPYSVSSONLKVSSLEQGFREFVSIIDHLDPDKYALHSHVYEGCAGSHVYL 202
DB 234 ROLGFESAHRTEVA-----HDFEANSFSLRYN--STIOESLR-RSHCPQORYISL 282
QY 203 OCTACGHRGYSRRIYGNMISLQPMQASLOFQYHLCGGSVITPLIITAHCYVDL 262
DB 283 QCSHGC-LRAMTGRIVGALASDKRPMQVSLHFGTHICGGLTIDAOVULRAHCFEYT 341
QY 263 --YLPKSWTIQVGLSLDNPAPSHLVEKIYVHSKKPRKGLNDIALMLKAGPLTFNEMI 320
DB 342 REKLEGMKVYAGTGNLHOLPEASIAE-IIINSNTDEEDDYDIALMLRLSKPLTSLAH 400
QY 321 QVVCPLPNSSENPFGKVCMTSGCATDGDGADSPVLNHAAPLISNKICNHRVYGGI 380
DB 401 HPACLPMMHQTPLNWCETITGFGKTRTDKTSPLREVONLIDFKCNDYLVYDYL 460
QY 381 SPMSLACAGYLTGVDSCQSGGLVCOERLMLKVGATSGFICAEVKNKPGVYTRVTSF 440
DB 461 TTRMNCAGDLHGGRDSCQSGGLVCOERLMLKVGATSGFICAEVKNKPGVYTRVTSF 520
QY 441 LPMIHEQMERDLK 453
DB 521 LPMIYSKMESEVR 533
RESULT 5
Q97306

ID 097506 PRELIMINARY: PRT: 643 AA.
AC 097506:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE KALLIKREIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLASMA;
RA Takahashi T., Kimura A., Okimura H., Hamabata T.;
RT "Porcine liver plasma kallikrein."
RL Submitted (JAN-1999) to the EMBL/Genbank/DBD databases.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: AB022425; BAA37147.1;
DR HSP: P00766; ICHG.
DR MEROPS: S01.212;
DR InterPro: IPR000177; Apple.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR003014; PAN.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00024; PAN; 4.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00005; APPLIEDOMAIN.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00223; APPLE; 4.
DR SMART: SM00020; TRYP-Spc; 1.
DR PROSITE: PS00495; APPLE; 4.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 643 AA: 72227 MW: AFE2923E3C3CB80A CRC64;
Query Match 23.7%; Score 579; DB 6; Length 643;
Best Local Similarity 36.9%; Pred. No. 7.4e-46;
Matches 128; Conservative 55; Mismatches 118; Indels 46; Gaps 10;
QY 132 DMKGHYANACQLGFPYSVSSDNLVSSLEQGFREFVSIID-----HLL 176
DB 309 DEGEELNV-----TFVGANLQGFECTIRQFTYLSHPDCKGCKSLRLS 360
QY 177 PDDKYALHSHVYVRG-----CASGHVYTLQCTACGHRGYSRRIYGNMISLQPMQ 231
DB 361 SDGSPFKIRHGMKASSGYSRLCRSG---DHSACATKA--NTRIVGDTDFLGEMPMQ 413
QY 232 ASLQFO---GYHLGGSVITPLIITAHCYVDLYLPKSMITQVGL--VSLDNPAPSHL 286
DB 414 VSLQAKLRQNHLCGGISLIGHQWVLAHACFGLSLPDIWRIYIGGLINISLTKETPEQ 473
QY 287 VERIYHYSKPKRRLNDIALMLKAGPLTFNEMTQVCLPNSSENPFGKVCMTSGCAT 346
DB 474 VKEITHQYKILLESCHDIALKLEPLRYTDFQKVICLPSRDTWVYTNCKVWTGFT 533
QY 347 EGCAGDASVYLNHAAPLISNKIC--NHRDVTGGIISPMILCAGYLTGVDSCQSGSGP 404
DB 534 EE-KETIQNIIQVNPILVSNECOKSYRD---HKSIMKICAGYEGGCKDACKGSGSP 589
QY 405 LYCOERLMLKVGATSGFICAEVKNKPGVYTRVTSFLDWIHQOMERD 451
DB 590 LVCKYNGIHLVGTTSWGECARQEPGYTVYIETMDILKETO 636
RESULT 6
Q9DB10 PRELIMINARY: PRT: 799 AA.
AC Q9DB10:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

```

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE 1300008A22RIK. PROTEIN.
GN 1300008A22RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wushaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC (LDLRA) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL: AK004939; BAB23684.1;
DR MGD: MGI:1919003; 1300008A22RIK.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00057; ldl_recept_a; 3.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PRO0722; CHYMOTRYPSIN.
DR PRINTS: PRO0261; LDLRECEPTOR.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00192; LDLa; 3.
DR SMART: SM00020; tryp_Spec; 1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS50068; LDLa_2; 3.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Glycoprotein; Hydrolase; Serine protease.
SQ SEQUENCE 799 AA; 89557 MW; 16315A64A5D288 CRC64;

```

```

DB 558 LOC15SRIVGTVSSEGEWPMOASLQIRGRHICGALLIADRWVITTAHCFQEDSMASPKL 617
QY 268 WTQVGVSLNDNP----APSHLYEKIVYSHKPKPRGLNDILMKLAGLFTFNEMIOPV 323
DB 618 WTFLG--KMRONSRRPGEVSEFVSRFLPHRYHEEDSHDYDALDQDHPVYSATVRPV 675
QY 324 CLPNSSENPDPGKWTSCGATEDGAGDASPLYNAAPPLISNKNRHYVGGIISPS 383
DB 676 CLPARSHFEPRGHCHMTWGAGREG--GPVSNLTQKVDVQVLPQDLS--EAVRYGVSPR 732
QY 384 MLCAGYLTGVDSCQDGSGLVCOE--RLMKLVGATSRGICGAENVKPEVYTRVTSFLD 442
DB 733 MLCAGYRKGRKACQDSDGSLVLCREPSGRMFLAGLVSMGLGCGRBNFEGVYTRVTVIN 792
QY 443 WHE 446
DB 793 WIDQ 796

```

RESULT 7
09NMSS PRELIMINARY: PRT: 1322 AA.

```

AC 09NMSS;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE SERINE PROTEASE 22D.
GN SP22D.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxId=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G3; TISSUE=HEMOLYMPH;
RX MEDLINE=20110889; PubMed=10646969;
RA Gorman M.J., Andreeva O.V., Paskewitz S.M.;
RT "Molecular characterization of five serine protease genes cloned from
RT Anopheles gambiae hemolymph.";
RT Insect Biochem. Mol. Biol. 30:35-46(2000).
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC (LDLRA) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL: AF117751; AAD38337.3;
DR InterPro: IPR002557; Chitin_binding.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR001190; SRCR.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF01607; Chitin_bind_2; 2.
DR Pfam: PF00057; ldl_recept_a; 2.
DR Pfam: PF00530; SRCR; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PRO0722; CHYMOTRYPSIN.
DR PRINTS: PRO0261; LDLRECEPTOR.
DR PRINTS: PRO0258; SPERACTRCPTR.
DR SMART: SM00494; CHBD2; 2.
DR SMART: SM00192; LDLa; 2.
DR SMART: SM00202; SR; 2.
DR PROSITE: PS01209; LDLa_1; 1.
DR PROSITE: PS50068; LDLa_2; 2.
DR PROSITE: PS00420; SRCR_1; UNKNOWN_1.
DR PROSITE: PS50287; SRCR_2; 2.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Glycoprotein; Hydrolase; Protease; Serine protease.
SQ SEQUENCE 1322 AA; 146811 MW; 2707110783A6B843 CRC64;

```

Query Match 23.2%; Score 567; DB 5; Length 1322;

Best Local Similarity 31.3%; Pred. No. 2,6e-44;
Matches 144; Conservative 68; Mismatches 174; Indels 74; Gaps 15;

```

QY 53 VITIIALLALALGIGIHFDCCSGKY-RCRSSFCIELIARCDVSPCKOEDEX--RC-- 107
Db 872 VGVGVCRTPVMS-----CPDYMWLCCHASECEIPVQFLCDNVRCACADSDSPHCKA 923
QY 108 ---VR-VGG---ONAVLOVFTASMKTCSDMKGHYANVACAQLGFPSSVSDNLRVSS 160
Db 924 PLAVRLVAGTDRGVEIINHGWTGTCDDDDGCVREARICQQLGFGNCAEVRKSYRPP 993
QY 161 LBSQFREEFVSIDHLIP--DDKYTALHSHVYREGCASHVTLQC----- 204
Db 984 GVGQIMLDVACNGTEPSIEDCV---HMHGHEHNCANHEDEYVRCGVYVTPKAPARLR 1039
QY 205 -----TAGC-----HRRGYSRTYVGNMSSLISQWPQASLQFP 237
Db 1040 ATRPNRPFDEVERSKIHDPDTCGRVILDPTRRPTYGARVHSGSEVYGHHPQASLRVK 1099
QY 238 GYHLCGSGVITPLMITTAACHVYDLYLPKS-WTIQVG--LVSLDNPAPSHLEKITYHS 294
Db 1100 TMMCGAVLITRHYVLTAAHCL--IGYPKSTYRVRIGDYHTAAYDNAELDIFIENTYIHE 1157
QY 295 KYRP-KRLGNDIALMKLAGELTFNEMIQPVCLPENSENEPDKVYWTSGMATEDAGADA 353
Db 1158 QREGHMMSNDIAYVVLKTFVRENDYVQPICLPAPADAPYLPQONCTISGMATEAGSKDS 1217
QY 354 SPLNHAAYPLISNKCINHDVYGGIISPMICAGYLITGVCQDSDGSGPLVC-QERRL 412
Db 1218 SYDLRAGVYPLLPDSCRRPEYVGDSDLIDGFCAGTLEPVDSCDDSGGLVPCPNSGL 1277
QY 413 WKLVGATSEFGICGAENVKPGVYTRVTSFLDMIHOMERDL 452
Db 1278 HTLTGIYSWKGKCGYANKPGVYLKVAHYRDMIOKLNQSL 1317

RESULT 8
ID 060235 PRELIMINARY; PRT; 418 AA.
AC 060235;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE AIRWAY TRYPSIN-LIKE PROTEASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98234382; PubMed=9565616;
RA Yamada K., Masuda K., Ogawa H., Takagi K., Umemoto N., Yasuoka S.,
RT "Cloning and characterization of the cDNA for human airway trypsin-
like protease."
RT J. Biol. Chem. 273:11895-11901(1998).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
Db EMBL: AB002134; BAA28691.1;
DR HSSP: P00750; 1RTF.
DR MEROPS: S01.301; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000082; SEA.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF01390; SEA; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPsin.
DR SMART: SM00200; SEA; 1.
DR SMART: SP00200; TRYP_Spc; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 418 AA; 46263 MW; F4RCIDB020CFBBD0 CRC64;

```

Query Match 23.1%; Score 564.5; DB 4; Length 418;
Best Local Similarity 34.2%; Pred. No. 9.7e-45;
Matches 142; Conservative 68; Mismatches 166; Indels 39; Gaps 13;

```

QY 51 IIVIGIALLALALGIGIHFDG--SGKYRCSSFCF--IELIARCDVSGCKDGEDYR 106
Db 22 IVVAGVV--ILAVTIALVYFLAFDQKSYFYRFSOLLANEVNSQLN-----SPATQEYR 74
QY 107 CVRVGQNAVLOVFTASMKTM-----CSDMKGHYANVACA-QLGFPSSVSDNLR 157
Db 75 TLSGRIESLTTKTFKESNLNPFIRAHVAKLRDQGSVRAADVWVKFQFTBRNNGASMKSR 134
QY 158 VSSLEGQFREEFVSIDHLIPDDKYTALHSHV---YVREGCASHVTLQCTACGHRGY 213
Db 135 IESVLAQMLNNSGNLE-INSTEITSLTDQAAANWMLINEGCAGPDLITLS----- 183
QY 214 SSRIVGQNMSSLISQWPQASLQFGYHLCGSGVITPLMITTAACHVYDLYLPKSWTIQVG 273
Db 184 EQRILGTGEAEESQSPVQVSLRLNNAHHCQSGSLINNMWILTAACHFRSNGSPRDMWITSG 243
QY 274 LVSLDNPAPSHLYEKIVYISKYKPKRLGNDIALMKLAGELTFNEMIQPVCLPENSENEFP 333
Db 244 ISTTF--PKLRMRVRNILLNNYKSAETHENDIALVRLNENSVETPKDIHVSCLPAAQONIP 301
QY 334 DGKVCWTSGMATEDAGDASPVLNHAAYPLISNKCINHDVYGGIISPMICAGYLITG 393
Db 302 PGSTAYVTYNGA-QETAGHIVPELRQGOVRIISNDVCNAPASHNGAILSCMLCAGVPQGG 360
QY 394 VDSQGDSDGSGPLVCQE-RRLMKLVGATSEFGICGAENVKPGVYTRVTSFLDMIHQ 447
Db 361 VDACQGSQSGPLVQEDSRRLMFLIVGIYSWGDQGLPDKPGVYTRVAYLDMIRQ 415

RESULT 9
ID 09NATO PRELIMINARY; PRT; 1322 AA.
AC 09NATO;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ADHESIVE SERINE PROTEASE.
GN SP22D.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20318993; PubMed=10860981;
RA Danilelli A., Loukeris T., Lagueux M., Mueller H.M., Richman A.,
RA Kafatos F.C.;
RT "A modular chitin-binding protease associated with hemocytes and
RT hemolymph in the mosquito Anopheles gambiae."
RT Proc. Natl. Acad. Sci. U.S.A. 97:7136-7141(2000).
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
(CC (LDLRA) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
Db EMBL: AJ276428; CAB81934.1; -.
DR InterPro: IPR002557; Chitin binding.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_recept_A.
DR InterPro: IPR001190; SSCR.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF01607; Chitin_bind_2; 2.
DR Pfam: PF00057; ldl_recept_a; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPsin.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR00258; SPERACTRCPTR.

```

DR SMART; SM00494; ChtBD2; 2.
 DR SMART; SM00192; LDla; 2.
 DR SMART; SM00202; SR; 2.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS01209; LDla_1; 1.
 DR PROSITE; PS0068; LDla_2; 2.
 DR PROSITE; PS00420; SRCR_1; UNKNOWN_1.
 DR PROSITE; PS50287; SRCR_2; 2.
 DR PROSITE; PS00134; TRYP_SIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYP_SIN_SER; 1.
 KW Glycoprotein; Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 1322 AA; 146794 MW; FBP973C21CC5475B CRC64;

Query Match 23.0%; Score 564; DB 5; Length 1322;
 Best Local Similarity 31.1%; Pred. No. 5, 1e-44;
 Matches 143; Conservative 68; Mismatches 175; Indels 74; Gaps 15;

QY 53 VIGIILALALGIGIHFDSCGY-RCRSSFCIELIARCDSVDCDEDEY--RC-- 107
 DB 872 VVGVCVCTPYMS-----CPDYMLCHASSECIPOVFLCDNVRDCADGSDSPDHCKA 923
 QY 108 ---VR-VGG---QNAVQVFTASMKTCSDMKGHYANACADLGFPSVSSDNLAVSS 160
 DB 924 PLAVRLVAGPTDREGREIVHGTGTCVDDFVREARVYICRQLGFNGTAEVAKSVYPP 983
 QY 161 LEGQFREFFSIDHLPR--DDKYALHHSYVREGCASGHVYTLQC----- 204
 DB 984 GVGQIMLDQVACNGTEPSIDCV--HMHGHEHNGCHTEDEYVRCGVYPTKRAARLR 1039
 QY 205 -----TACG-----HRRGYSRIYGVGMSLISQMPWQASLQFO 237
 DB 1040 ATRPNRPFDEVERSRIKIHPTGCGVLLIDPLRLKPTYGARVHGSEYVGHHPQASLRK 1099
 QY 238 GYHLCGGSVTPMLITIAHCVYDLRLKS--WTLQVG--LVSLDNPAPSHLVKIIYHS 294
 DB 1100 TMMHCGAVLITRHHVLAACHL--IGYKSTYRIRIGDYHTPAAYDNLDFIENTYIHE 1157
 QY 295 KYRP-KRLGNDIALMKLAGPLTFENMIQVCLPNSSENFPPDGKVCWTSFGATEDEGAGA 353
 DB 1158 QFREGHMSNDIAVVLVKTFRFNDYQVPCILPARADAPYLPQGCSTISMGATFAGSKDS 1217
 QY 354 SPVLNHAAYPLISKICNHDVYGGIISPSMLCAGYLTGVDSCGDSGGPILVC-QERRL 412
 DB 1218 SYDLRACTVPLRPSVCRREYVGDLSLIDCMFAGLTPEPVDSCDSDGSPILVCPSNEGL 1277
 QY 413 WKLVSFSGICAEVKNPGVYTRVTSFLDMIHQEMERDL 452
 DB 1278 HTLTGIYSMGKHCYANKPGVYTLKVANHRTMIEKLNQSL 1317

RESULT 10
 Q9Y1V3 PRELIMINARY; PRT; 868 AA.
 AC Q9Y1V3; 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
 DE TUNICATE RETINOIC ACID-INDUCIBLE MODULAR PROTEASE PRECURSOR.
 GN TRAMP.
 OS Polyandrocampa misakiensis.
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
 OC Stolidobranchia; Styelidae; Polyandrocampa.
 OX NCBI_Taxid=7723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WHITE SPOT.
 RA Ohashi M., Kawamura K., Fujii N., Yubisui T., Fujiwara S.;
 RT "A retinoic acid-inducible modular protease in budding ascidians."; Dev. Biol. 0:0-0(1999).
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A (LDLR) DOMAIN
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.

CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL; AB030007; BAA82522.1; -
 DR HSSP; P00750; IRTF.
 DR Interpro; IPR001314; Chymotrypsin.
 DR Interpro; IPR002172; LDL_recept_A.
 DR Interpro; IPR003609; Pan_app.
 DR Interpro; IPR001254; Trypsin.
 DR Pfam; PF00057; 1dl_recept_a; 3.
 DR Pfam; PF00530; SRCR; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRINTS; PR00258; SPERACTRCPTR.
 DR SMART; SM00192; LDla; 3.
 DR SMART; SM00473; PAN_AP; 1.
 DR SMART; SM00202; SR; 2.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS01209; LDla_1; 3.
 DR PROSITE; PS50068; LDla_2; 3.
 DR PROSITE; PS50287; SRCR_2; 2.
 DR PROSITE; PS00134; TRYP_SIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYP_SIN_SER; 1.
 KW Glycoprotein; Hydrolase; Protease; Serine protease; Signal.
 FT SIGNAL 1
 SQ SEQUENCE 868 AA; 97659 MW; F7146285F36ACA CRC64;

Query Match 22.4%; Score 548.5; DB 5; Length 868;
 Best Local Similarity 32.3%; Pred. No. 8, 3e-43;
 Matches 146; Conservative 59; Mismatches 148; Indels 99; Gaps 19;

QY 72 DC--SGYRRSSPFCIELIARCDSVDCDEDEYRC-----VVGQNAV-----LQVFT 121
 DB 434 DCATNNYILCNDG--SCLEHQQVCFRDPCNGEETCECEPMKIRLRGSGPFECHYEVAK 492
 QY 122 AASKTKMSDDWKGHYANVACAQLGF-----PSYSSDNLRYSSLEGO 164
 DB 493 GRKYGILCDTRMISIREADVVCRLQGFERRGAIIDALKGAGFRIDRRV--LDNVRCNG---- 547
 QY 165 FREEFVSIDHLPRDPDKYALHHSYVREGCASGHVYTLQC----- 204
 DB 548 -RE--SLEN-----CRHSGMKNKASCSDIEHGVICRQEATTPSPSPSATPQPTP 595
 QY 205 -----TACGHRGY-----SSRIYGVMSLISQMPWQASLQFOGYHLCGGSVIT 248
 DB 596 RRIITTPRPMPREGKRPVLEAPLTARIYGSSTTEHHPMGAGIVLPPTWYMGSLIH 655
 QY 249 PLWITTAHACVYDLYLPKSWTIQVG--LVSLDNPAPSHLVEKIVHSHKYPKRLGNDIA 306
 DB 656 PCWVLTAAHCFVREYPRIDYIRLGDHITGVDETEOLFRIAEIHKHD--YNVTTKENDIA 714
 QY 307 LMKLAGP-----LTFENMIQVCLPNSSENFPPDGKVCWTSFGMGTEDGAGA--SPVLNHA 361
 DB 715 LRLTENDARCATTPEVQVCLPKSSQFPDAKTICVETGWDGSDATVAVYVPLEAE 774
 QY 362 VPLISNKCINHRDVGGLIISPSMLCAGYLTGVDSCGDSGGPILVCQERR-----LMKLV 416
 DB 775 IPLIANKKC-LRSEVYQLQPTMFCAGYLTGVDSCGDSGGPILVCQSDDDRYVW--- 830
 QY 417 GATSFSGICAEVKNPGVYTRVTSFLDMIHQEM 448
 DB 831 GIVSWGKCAKPKAPGAYKAVAFIDWI--EQM 861

RESULT 11
 Q9JUT7 PRELIMINARY; PRT; 855 AA.
 AC Q9JUT7; 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE MEMBRANE BOUND SERINE PROTEASE (MEMBRANE BOUND ARGININE SPECIFIC


```

DE SERINE PROTEASE).
GN MBSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=JEJUNUM;
RA Tsuzuki S.;
RT "A membrane bound serine protease expressed in rat small intestine.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=DUDENUM;
RA Inoue H., Takahashi K., Kishi K.;
RT "membrane-bound arginine specific serine protease.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
(CC (LDLR) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR EMBL: AB037898; BAB03502.1; -
DR EMBL: AB049188; BAB13765.1; -
DR InterPro: IPR002106; AA_trna_ligase_II.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR002172; LDL_recept_A.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00057; ldl_recept_a; 4.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00192; LDL; 3.
DR SMART: SM00020; tryp-spec; 1.
DR PROSITE: PS00339; AA-trna_ligase_II.2; UNKNOWN_1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS50068; LDLRA_2; 4.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR GlycoProtein: Hydrolase; Protease; Serine protease.
FT VARIANT 665 K->N.
FT SEQUENCE 855 AA; 94955 MW; 35806b7ec6c03d CRC64;

```

```

Query Match 22.2% Score 542.5; DB 11; Length 855;
Best Local Similarity 31.5%; Pred. No. 3e-42;
Matches 135; Conservative 63; Mismatches 148; Indels 83; Gaps 16;

```

73 CGGKRCRCSFICILIANCDGVSCCKDEDEYRCYRVGGQNAVLQVFAAASKTKMCSND 132
 453 CGGMEFCKTG-CIKRDLFCDCWADCPDSDERHC-----RCNATHQFCXKXNO 499
 133 -----WKGHYANVACAQLGFPYSVSSDNLRYSLLEGQREFEVSIDHLLPD----- 178
 500 FCKPLFWDDSVN-DCGD-----GSDEGSCSPAGSK-----CSNGKLLPSSQCCNGKD 548
 179 -----DKVTALHHSVYRREGCASGHVVTLLQCTACAGRGYS----- 214
 549 DCGDGSDEASCNVNAVSCSTKYTYR-CQNGCLCNKNGNPECDCGKCCDSDEKNCDCGAR 607
 215 -----SRIYGMNLSLQMPQASLQFGY-HLCGGSVITPLMITTAACVVDILPK-- 266
 608 SFTKQARRVGVGNADGEMPMQVSLHALGQHLGCGASLSPWLVSAAHCFODETIFKYS 667
 267 ----SWTIQGLVSLDNPV----PSHLVERKIVYHSKKPKRLNDLALMKIAPLTFENMI 320
 668 DHTMTAFGLGLDQSKRSASGVQEHKRLRIITHPSFNDTTPYDIALLEKPAEISTYV 727
 321 QPVCLPNSSENFPPDGKVCWTSGMGATGEGAGDASPVLAHAAPVLLISNKICNHRDVGII 380

```

DB 728 RPICLPDNTHVFPAGKAIWVTGWHTRKEG-GTGALLIQGELIVINQTC--ELLPPQI 784
QY 381 SPSMCAAGLTGTVSCDSCDGGSPGLVCOQR--RLMKLVGATSGICGCAVNPQVYTRT 438
DB 785 TPRMKCVGFLSGVSDSCDGGSPGLSSVERKGRIFQ-AGVWSGEGCAORNNPQVYTRIP 843
QY 439 SFLDWIHEQ 447
DB 844 EYRDWIKED 852

```

RESULT 12
 Q9VSU2 PRELIMINARY; PRT; 1186 AA.
 AC Q9VSU2: Q9VSU1;
 DT 01-MAR-2000 (TrEMBLrel. 13, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE TEQUILA PROTEIN.
 GN TEQUILA OR CG4821 OR CG4948 OR CG18403.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OX Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brocktein P., Brotlier P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jajall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spiter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
 (LDLR) DOMAIN.
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL: AE003553; AAF50319.2; -

DR HSSP; P20231; 1AAO.
 DR FlyBase; FBgn0023479; Tegulla.
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_recept_A.
 DR InterPro; IPR002965; P_rich_extensions.
 DR InterPro; IPR001190; SRCR.
 DR InterPro; IPR001254; trypsin.
 DR Pfam; PF00057; Idl_recept_a; 2.
 DR Pfam; PF00530; SRCR; 2.
 DR Pfam; PF00089; trypsin; 2.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRINTS; PR01217; PRICHEXTENSN.
 DR PRINTS; PR0258; SPERACTRCPTR.
 DR SMART; SM00192; LDla; 2.
 DR SMART; SM00202; SR; 2.
 DR SMART; SM00020; TRYPSIN; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
 DR PROSITE; PS01209; LDla_1; 1.
 DR PROSITE; PS50068; LDla_2; 2.
 DR PROSITE; PS00420; SRCR_1; UNKNOWN_2.
 DR PROSITE; PS50287; SRCR_2; 2.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR GlycoProtein; Hydrolase; Serine protease.
 KW Glycoprotein; 1186 AA; 131246 MW; 75474D3968BA32D CRC64;
 SQ

Query Match 22.0%; Score 539; DB 5; Length 1186;
 Best Local Similarity 27.5%; Pred. No. 9.9e-42;
 Matches 134; Conservative 74; Mismatches 167; Indels 112; Gaps 15;

QY 66 GLGIH-----FDSCGKY-RCRSSFCIEILACDGVSDCKDEDEYRCV 108
 DB 709 GMEVHNGVDVAGVYCKVYVMKCPNNYWLCHTSKECIPAFVCDNPPDADKSD--CA 766
 QY 109 RV-----GGQNA---VLOVFTASWTKMCSDDMKGHYANVACAQLFPSSVSSDN 155
 DB 767 AVCAQAPQVRLLEGRLNENGRLEVKHGVGSCVDDDFNLKSAQVACNSMGFGPAKIEK 826
 QY 156 LRVSSLEGQREFEVSIDHL---PDDKYALHHSYVVRGCSAGHYVTLQCTA----- 206
 DB 827 NIFGNSNGP-----IWLDOVCFGENETSIDQCNHMMNGEHNCHTEDVALHCSAGPPRS 881
 QY 207 -----CG-----HRGYS 214
 DB 882 QRSQIQIKGGRSLGRFTTPKITYSQIGLMERSKAVHTPRRCGLFEKDDLDEVAHR---E 938
 QY 215 SRIYVGNMSSLQMPWQASIQFOG-----YHLCGGSVITPLMTITTAHCYVDLYLPK-SW 268
 DB 939 ERYVVRGVNAQRGHRPMQATIRTRGRGSISSHMGCAVVISKRHLTAHCLYG--SPKGA 996
 QY 269 TIQVG--LVSLLDNPASHLEKIVYHSKY-KPKRLGNDIALMKLAGPLTFNEMIQVCL 325
 DB 997 FVRVGDHYANIAASSEVDSTENMYLHENFRKGTTHMNDIALVVLKTPLEKFSVDQICL 1056
 QY 326 PNSEENFPPDKVCTSGMGATPDGADSPVLNHAAPVLSNKCINRDVYGGITISML 385
 DB 1057 PDKNAELVEDRKCTISGWSIKSGVSTPPAUVLSAELPIIADHYCKOSNYYGSMSSGMF 1116
 QY 386 CAGLTGCVDSQGDGSGPLVCOERRMLKLVGATSFSGICAEVKKPQVYTVTSFLDMIH 445
 DB 1117 CAGSMESVYACEDSGDGLVCSDDDEFTLGLISWGHCGFKNRPGVYVYVNHVIMIX 1176
 QY 446 EOMERDL 452
 DB 1177 EKINESL 1183

RESULT 13
 090112 PRELIMINARY; PRT; 1449 AA.

AC Q90112;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE GRAAL PROTEIN.
 GN TEQUILLA OR GRAAL OR CG4821 OR CG4948 OR CG18403.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Munier A.I., Medzhitov R., Janeway C., Hoffmann J.A., Laguex M.;
 RT "Characterization of a new serine protease in Drosophila
 RT melanogaster";
 RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
 CC (LDLRA) DOMAIN.
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL; A1251803; CAB64653.1; -.
 DR HSSP; P20231; 1AAO.
 DR FlyBase; FBgn0023479; Tegulla.
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR InterPro; IPR002557; Chitin_binding.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_recept_A.
 DR InterPro; IPR001190; SRCR.
 DR InterPro; IPR001254; trypsin.
 DR Pfam; PF01607; Chitin_bind_2; 2.
 DR Pfam; PF00057; Idl_recept_a; 2.
 DR Pfam; PF00530; SRCR; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRINTS; PR0258; SPERACTRCPTR.
 DR SMART; SM00494; CtlBD2; 2.
 DR SMART; SM00192; LDla; 2.
 DR SMART; SM00202; SR; 2.
 DR SMART; SM00020; TRYPSIN; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
 DR PROSITE; PS01209; LDla_1; 1.
 DR PROSITE; PS50068; LDla_2; 2.
 DR PROSITE; PS00420; SRCR_1; UNKNOWN_2.
 DR PROSITE; PS50287; SRCR_2; 2.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR GlycoProtein; Hydrolase; Serine protease.
 KW Glycoprotein; 1449 AA; 160045 MW; F3BC806543CAB6ED CRC64;
 SQ

Query Match 22.0%; Score 539; DB 5; Length 1449;
 Best Local Similarity 27.5%; Pred. No. 1.3e-41;
 Matches 134; Conservative 74; Mismatches 167; Indels 112; Gaps 15;

QY 66 GLGIH-----FDSCGKY-RCRSSFCIEILACDGVSDCKDEDEYRCV 108
 DB 972 GMEVHNGVDVAGVYCKVYVMKCPNNYWLCHTSKECIPAFVCDNPPDADKSD--CA 1029
 QY 109 RV-----GGQNA---VLOVFTASWTKMCSDDMKGHYANVACAQLFPSSVSSDN 155
 DB 1030 AVCAQAPQVRLLEGRLNENGRLEVKHGVGSCVDDDFNLKSAQVACNSMGFGPAKIEK 1089
 QY 156 LRVSSLEGQREFEVSIDHL---PDDKYALHHSYVVRGCSAGHYVTLQCTA----- 206
 DB 1090 NIFGNSNGP-----IWLDOVCFGENETSIDQCNHMMNGEHNCHTEDVALHCSAGPPRS 1144
 QY 207 -----CG-----HRGYS 214
 DB 1145 QRSQIQIKGGRSLGRFTTPKITYSQIGLMERSKAVHTPRRCGLFEKDDLDEVAHR---E 1201
 QY 215 SRIYVGNMSSLQMPWQASIQFOG-----YHLCGGSVITPLMTITTAHCYVDLYLPK-SW 268

```

Db 1202 ERYVRNVAQRGRHPQATIRTRGCGISSHMGAVYISKRHLITFAHCLYG--SPKGAY 1259
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 269 TIOVG--LVSLDNPAPSHLVEKIVYHSKY-KKRLGNDIALMKLAGPLTFNEMIQPVCL 325
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1260 FVAVGHHYANIAESVDSFIENWYLHENRKTGHNMNDIALVLTPLKFSYVOPICL 1319
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 326 PNSEENFPDGKVCWTSGWGATEDEGADGADSPVNLHAAVPLISNKCINHRDVGIIISPSML 385
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1320 PDKNAELVEDRCKTIGSGKSGVSTPAVYLSAELPLIADHYVCQSNVYGSAMEEGMF 1379
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 386 CAGYLTGVDSGCGSGGGLVCGERLMLKLVGATSFSGICAEVKNKPGVYTRVTSFLDWIH 445
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1380 CAGSMDESVDACGSDGGLVCSDDDGELTYGLISWGHCGRKRRPGVYRVNHYIDWIT 1439
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 446 EQMERDL 452
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1440 EKINESL 1446
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::

```

```

RESULT 14
Q9B119 PRELIMINARY; PRT: 1462 AA.
AC 090113;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE GRAL PROTEIN.
GN TEOULTA OR GRAL OR CG4821 OR CG4948 OR CG18403.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Munier A.I., Medzhitov R., Janeway C., Hoffmann J.A., Lagueneux M.;
RT "Characterization of a new serine protease in Drosophila
RT melanogaster.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC (LDLR) DOMAIN.
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (SI).
CC EMBL: AJ251802; CAB64652.1; -.
DR HSSP: P00750; ITRF.
DR FLYbase: FBgn0023479; Tegula.
DR InterPro: IPR000194; Atpase_alpha_beta.
DR InterPro: IPR002557; Chitin_binding.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_recept.A.
DR InterPro: IPR002965; P_rich_extenso.
DR InterPro: IPR001190; SRCL.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF01607; Chitin_bind_2; 2.
DR Pfam: PF00530; SRCL; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00261; LDLRECEPTOR.
DR PRINTS: PR01217; PRICHTEXTEN.
DR PRINTS: PR00258; SPERACTRCPTR.
DR SMART: SM00494; ChIBD2; 2.
DR SMART: SM00192; LDLa; 2.
DR SMART: SM00202; SR; 2.
DR SMART: SM00020; tryp_Spc1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS00068; LDLRA_2; 2.
DR PROSITE: PS00420; SRCL_1; UNKNOWN_2.
DR PROSITE: PS00287; SRCL_2; 2.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.

```

KW Glycoprotein; Hydrolase; Serine protease.
SQ SEQUENCE 1462 AA; 161444 MW; E3AC494FE140F37A CRC64;

Query Match 22.0%; Score 539; DB 5; Length 1462;
Best Local Similarity 27.5%; Pred. No. 1.3e-41;
Matches 134; Conservative 74; Mismatches 167; Indels 112; Gaps 15;

```

Qy 66 GIGIH-----FDCSGKY-RCRSFICIELIACDGVSOCKDEDEYRCY 108
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 985 GWGVHNCGVDEAVGTYCKVPMKCPNNYWLCHTSKCTIPPAFYCDNTPPCDADSDE--CA 1042
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 109 RV-----GQNA---VLQVTAASWKTQCSVDMMKHYANVCAOLFSPYSVDN 155
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1043 AVCGAIVQYRLLEGGRNSNEGRLEVKHHGWSYCDDDFNLSQVACNSMGFFGPAKIER 1102
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 156 LRVSLGQFREEFVSIDHL--PDKYVTAHHSVYVREGASGHVYTLQCTA----- 206
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1103 NIFGNSNGP-----IWLQVWCFGNFTSIDQNHMMWGEHNCNHTEDVALHCSAGPPRS 1157
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 207 -----CG-----HRGYS 214
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1158 QRYSTQIKGRSLGRTTPKTYSQIGLWERSKAVHTPRRCGIFKDDLTDEYAHN---E 1214
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 215 SRIVGNSMLISQWPQASLOFQG-----YHLCGSAVTPPLMITTAACHVYDLYLPK-SW 268
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1215 ERYVRNVAQRGRHPQATIRTRGCGISSHMGAVYISKRHLITFAHCLYG--SPKGAY 1272
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 269 TIOVG--LVSLDNPAPSHLVEKIVYHSKY-KKRLGNDIALMKLAGPLTFNEMIQPVCL 325
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1273 FVAVGHHYANIAESVDSFIENWYLHENRKTGHNMNDIALVLTPLKFSYVOPICL 1332
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 326 PNSEENFPDGKVCWTSGWGATEDEGADGADSPVNLHAAVPLISNKCINHRDVGIIISPSML 385
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1333 PDKNAELVEDRCKTIGSGKSGVSTPAVYLSAELPLIADHYVCQSNVYGSAMEEGMF 1392
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 386 CAGYLTGVDSGCGSGGGLVCGERLMLKLVGATSFSGICAEVKNKPGVYTRVTSFLDWIH 445
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1393 CAGSMDESVDACGSDGGLVCSDDDGELTYGLISWGHCGRKRRPGVYRVNHYIDWIT 1452
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Qy 446 EQMERDL 452
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::
Db 1453 EKINESL 1459
      :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|::: :|:::

```

```

RESULT 15
Q9B119 PRELIMINARY; PRT: 2382 AA.
AC 09B119;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE GRAL2 PROTEIN PRECURSOR.
GN GRAL.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Munier A.I., Medzhitov R., Janeway C.A., Lanot R., Zachary D.,
RA Capovilla M., Lagueneux M.;
RT "Gral a Drosophila gene coding for several mosaic serine proteases.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ309005; CAC35209.1; -.
DR Signal.
FT SIGNAL.
SQ SEQUENCE 2382 AA; 264348 MW; 51C85282B0683D4 CRC64;

```

Query Match 22.0%; Score 539; DB 5; Length 2382;
Best Local Similarity 27.5%; Pred. No. 2.5e-41;

Matches 134; Conservative 74; Mismatches 167; Indels 112; Gaps 15;

```
OY 66 GLGIH-----FDGSGKY-RCRSSFKCIELIARCDGVSDDCKGDEDEYRCV 108
    |::|
Db 1905 GWCYHNGCVDEVAGVYTKVPMKCPNNYWLCHTSKECIPAFVCDNTPDCADKSD--CA 1962
OY 109 RV-----GGQNA--VLOVYTAASWKTMSDDMKGHYANVACAQLGFPSPYSSDN 155
    |::|
Db 1963 AVCOAPQVYRLGGRNSNEGRLLEVKHGHWGVCDDDFNLKSAQVACNSMGCFGPARIK 2022
OY 156 LRVSSLEGOFREEFVSIIDLH---PDDKYVTAHHSYVREGCASGHVYTLQCTA----- 206
    |::|
Db 2023 NIFGNSNGP-----IWLDOYWCFGNETSIDQCNMNMGEHNCNHTEDVALHCSAGPPPRS 2077
OY 207 -----CG-----HRRGYS 214
    |::|
Db 2078 QRSQTOIKGGRSLGRRTTPTKYSQIGLMERSKAAVHTPRRCGTFKDDLTDEYAHR---E 2134
OY 215 SRIVGMMSLISQMPQASLOFOG----YHLCGSVITPLMITTAHCVYDLYLPK-SW 268
    |::|
Db 2135 ERVVRGNVAORGRHPQATITRTGRGCISSHMCQAVVYSKRHLTLTAHCLYG--SPKGAY 2192
OY 269 TIQVG--LVSLDNPAPASHLEKIVYHSKY-KPKRLGNDIATLAKPLTFNEMIQPVCL 325
    |::|
Db 2193 FVRVGDHYANIAESSEVDSEFTENNYLHENFRKGTMMNDIALVVLKTPLEFSDYVOPICL 2252
OY 326 PNSEENPPDGKVCWTSGGATEDEGADASPYLNHAAVPLISNKCINHRDVGIIISPSML 385
    |::|
Db 2253 PDKNAELVEDRKCTISGWSIKSGSVTPPAQVYLSAELPILADHYCKOSNYGSAWSEGMF 2312
OY 386 CAGYLTGGVDSGCGDSSGPLYVCOERRLMKLVGATSFSGICAEVNNPGVYTRVTSFLDMIH 445
    |::|
Db 2313 CAGSMDESVDACEGDSGPLYVCSDDDEFTLYGLISMCGHCGFKNRPVYVYVNHVYDMY 2372
OY 446 EOMERDL 452
    |::|
Db 2373 EKINESL 2379
```

Search completed: April 11, 2002, 08:54:33
Job time: 121 sec

Db 652 GLGIPCECKEDNQCKDG-ECIPLVNLCDFPHCKDGSDEAHCVRLFNGTSTDSSGLVQFR 710

F:785-1014/Domain: trypsin homology <TRY>
E:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site

F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 26.8%; Score 657; DB 1; Length 1019;
Best Local Similarity 37.7%; Pred. No. 1.7e-44;
Matches 148; Conservative 67; Mismatches 148; Indels 30; Gaps 15;

67 LGIHFDGSGKRYCRSSFKCIELIARCDVSDCKDEDEYRCVR---VGQNAVLYQVPT 121
Db LGIPECKADHQCCKNG-ECVPLVNLCDHLEHEDSDADCVRFNGNTNNGLYRFRT 695
122 AASWKTMSDDMKGHYANVACAQGFPPSYSSDNLRVSSLEQGFREFFYSIDHLLPDDKY 181
Db OSIWHTACAEENWTTQISNDVQLGLGSGNSK--PIFSTDG--GPVKL-NTADG-- 747
182 TALHSHVYV-REGCASHVYVLTCT--ACGHR---RGYSRIYGGNMSLLSOWPWASLOF 236
Db ---HLLTPSQQLDLSRLQCNHKSCKKLAAQDITRKIVGSGNAKGAAMPVWVGLY 804
237 OGTHLCGGSVITPLMITTAHACVYDLYL-PSKWTIOVGL--VSLLDNP-APSHLVEKTY 292
Db GGLLGCASLVSMDLVSAHCVYGNLRPSKWTALGLHMSNLTFSPQTVRLIDELV 864
293 HSKYKPKRLGNDIALMKLAGPLTFENEMIOVCLPNSSENPDPGKVCMTSGWATEDGAD 352
Db NPHYNRRKNDNDIAMHLEFKNVYTDYIOPICLPENOVFPFGKNSIAGWG-TVYVQGT 923
353 ASFPVNLHAAVPLISNKCINHR-DVYGGIISPSMLCAGYLTGVDSCGDSGGLVCOERR 411
Db * 924 TAILIOEADVPPLISNEKCOQOMPEYN--ITENMCAGYEEGIDSCGDSGGLMCOENN 981
412 LMKLVGATSFEGICAEVKNKGVYTRVTSFLDVI 444
Db 982 RNFAGVTSFGYKCALPDRPGVAYARVSRTEWI 1014

RESULT 4

A53663
N:Enteropeptidase (EC 3.4.21.9) precursor - pig
M:Alternate names: enterokinase
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999
C:Accession: A53663
R:Matsumura, M.; Ichinose, M.; Yahagi, N.; Kakel, N.; Tsukada, S.; Mikl, K.; Kurokawa, J. Biol. Chem. 269, 19976-19982, 1994
A:Title: Structural characterization of porcine enteropeptidase.
A:Reference number: A53663; MUID:94327548
A:Accession: A53663
A:Molecule type: mRNA
A:Residues: 1-1034 <MAT>
A:Cross-references: GB:D30799; NID:9505122; PIDN:BA06459.1; PID:9505123
A:Note: parts of this sequence, including the amino ends of three chains isolated from t
C:Comment: The mechanism of association with the membrane of the intestinal brush border
C:Keywords: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding ref
C:Keywords: glycoprotein; hydrolyase; serine proteinase; transmembrane protein; zymogen
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MC>
F:118-799/Product: enteropeptidase heavy chain #status predicted <HC>
F:1199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:357-519/Domain: MAM homology <MAM>
F:341-646/Domain: C1r/C1s repeat homology <C1R>
F:658-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
F:800-1034/Product: enteropeptidase light chain #status predicted <LC>
F:800-1039/Domain: trypsin homology <TRY>
F:116,147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,96

F:787-911,825-841,925-992,956-971,982-1010/Disulfide bonds: #status predicted
F:840,891,986/Active site: His, Asp, Ser #status predicted

Query Match 26.7%; Score 654; DB 1; Length 1034;
Best Local Similarity 37.7%; Pred. No. 3e-44;
Matches 149; Conservative 65; Mismatches 147; Indels 34; Gaps 15;

67 LGIHFDGSGKRYCRSSFKCIELIARCDVSDCKDEDEYRCVR---VGQNAVLYQVPT 121
Db LGIPECKADHQCCKNG-ECVPLVNLCDHLEHEDSDADCVRFNGNTNNGLYRFRT 710
122 AASWKTMSDDMKGHYANVACAQGFPPSYSSDNLRVSSLEQGFREFFYSIDHLLPDDKY 181
Db OSIWHTACAEENWTTQISNDVQLGLGSGNSK--PIFSTDG--GPVKL-NTADG-- 757
182 TALHSHVYV-REGCASHVYVLTCT--ACGHR---RGYSRIYGGNMSLLSOWPWASLOF 234
Db TAPNSGLILTASQCEDESLILQCNHKSCKKQVQAEVSPKIVGNDREGAMPVWV 817
235 OGTHLCGGSVITPLMITTAHACVYDLYL-PSKWTIOVGL--VSLLDNP-APSHLVEKTY 290
Db YINGQLLCASLVSMDLVSAHCVYGNLRPSKWTALGLHMSNLTFSPQTVRLIDELV 877
291 VHSKYKPKRLGNDIALMKLAGPLTFENEMIOVCLPNSSENPDPGKVCMTSGWATEDGA 350
Db VINPHNRRKNDNDIAMHLEFKNVYTDYIOPICLPENOVFPFGKNSIAGWG-TVYVQGT 936
351 ASFPVNLHAAVPLISNKCINHR-DVYGGIISPSMLCAGYLTGVDSCGDSGGLVCOERR 409
Db 937 GSPADILQEADVPPLISNEKCOQOMPEYN--ITENMCAGYEEGIDSCGDSGGLMCOENN 994
410 RMLKLVGATSFEGICAEVKNKGVYTRVTSFLDVI 444
Db 995 NRMWLAGVTSFGYKCALPDRPGVAYARVSRTEWI 1029

RESULT 5

S33777
hepsin (EC 3.4.21.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C:Accession: S33777; S32013
R:Farley, D.; Raymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
A:Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinas
A:Reference number: S33777; MUID:93305733
A:Accession: S33777
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <FAR>
A:Cross-references: EMBL:X70900; NID:957928; PIDN:CAA50256.1; PID:957929
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:162-399/Domain: transmembrane #status predicted <TM>
F:187-203,250-358,321-337,348-380/Disulfide bonds: #status predicted
F:202,256,332/Active site: His, Asp, Ser #status predicted

Query Match 26.4%; Score 645; DB 1; Length 416;
Best Local Similarity 35.2%; Pred. No. 5.5e-44;
Matches 146; Conservative 61; Mismatches 156; Indels 52; Gaps 11;

60 IIALAIGLGHFDGSGKRYCRSSFKCIELIARCDVSDCKDEDEYRCVYRGQNAVLYQV 119
Db 17 VALATVGTLLPLTGIG---AASWALVTILR-----SDQEPYQVQLSPGDSRLVLV 65
120 -PTAASWKTMSDDMKGHYANVACAQGFPPSYSSDNLRVSSLEQGFREFFYSID----- 173
Db 66 DTEGTGTRLLCSRSNARVAGIGCEBEGFLRALHSELVVRTAGANGTSGFCVDEGGGLP 125
174 ---HLLPDDKVTALHSHVYVREGCASHVYVLTCTACGHRRGYSRIYGGNMSLLSOWPW 230

Db 126 LAORLL--DIVISVC-----DPRGRFLATCCODCGRRKRLPDRIVGCGDSSSLGRWP 175
 QY 231 QASLOFGYHLCGGSVITPLMITTAHCVYDLVLPK-----SWTIOGVLSLDNAPS 284
 Db 176 QVSLRDGTHLCCGSLISGDMVLAHC-----PPENNRSLRRKRVAGAVARTSPHAYQ 230
 QY 285 HLEKTIYHSHKYPKR-----LGNDIALMKLAGPLTFENMIOPVCLPNSSENPDKVC 338
 Db 231 LGQAVAVITHGGYLPFRNPPTIDENSNDIALVHSSLPLEIYQPCVPAGQALVDSKVC 290
 QY 339 WISGWCATEDGADASPVYLNHAAPLISNKCINHRDVGIIISPSMTCAGYLTGVDSCQ 398
 Db 291 TVTWGWTQ--FYGQAVLVLOEARPIISNEVCNSPDYGNQIKPKMFCAGYPEGIDACQ 349
 QY 399 GDSGGPLVCOER-----RLMKLVGATSFICGCAEVNKPQVYTRVTSFLDWMHEOME 449
 Db 350 GDSGGHEVCEDRLSGTSGRMLCGIVSWGTCALARKPGVYTKVYIDFREMIFQAIK 404

RESULT 6 KORTPL

Plasma kallikrein (EC 3.4.21.34) precursor - rat
 N:Alternate names: Fletcher factor; kininogenin; serum kallikrein
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
 C:Accession: A39180; A33320; S06851; I53041; S06852
 R:Beaubien, G.; Rosinski-Chupin, I.; Mattei, M.G.; Mokay, M.; Chretien, M.; Seidah, N.G.
 Biochemistry 30, 1628-1635, 1991
 A>Title: Gene structure and chromosomal localization of plasma kallikrein.
 A:Reference number: A39180; MUID:91129236
 A:Accession: A39180
 A:Molecule type: DNA
 A:Residues: 1-638 <BEA>
 A:Cross-references: GB:J05315
 A>Note: the authors translated the codon GAG for residue 81 as Glu
 R:Seidah, N.G.; Ladeheim, R.; Mokay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazut
 DNA 8, 563-574, 1989
 A>Title: The cDNA structure of rat plasma kallikrein.
 A:Reference number: A33320; MUID:90091743
 A:Accession: A33320
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-638 <SEI>
 A:Cross-references: GB:M30282; NID:q205010; PIDN:AAA41463.1; PID:q205011
 A>Note: part of this sequence, including the amino ends of both the heavy and light chain
 R:Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
 Biochim. Biophys. Acta 999, 103-110, 1989
 A>Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
 A:Reference number: S06851; MUID:90089457
 A:Accession: S06851
 A:Molecule type: protein
 A:Residues: 20-45;391-413 <PAQ>
 R:Seidah, N.G.; Ladeheim, R.; Mokay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazut
 DNA Cell Biol. 8, 563-574, 1989
 A>Title: The cDNA structure of rat plasma kallikrein.
 A:Reference number: I53041
 A:Accession: I53041
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-638 <RES>
 A:Cross-references: GB:M58590; NID:q206721; PIDN:AAA42069.1; PID:q206722
 C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
 are linked by one or more disulfide bonds.
 C:Genetics:
 A:Gene: PK
 C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; fibrinolytic; glycoprotein; hydrolase; infla
 F:1-19/Domain: signal sequence #status predicted <Sig>
 F:20-390/Product: plasma kallikrein heavy chain #status experimental <Matl>
 F:20-109/Domain: apple repeat <AP1>
 F:110-199/Domain: apple repeat <AP2>

F:200-289/Domain: apple repeat <AP3>
 F:291-380/Domain: apple repeat <AP4>
 F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
 F:391-621/Domain: trypsin homology <TRY>
 F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,
 F:127,215,308,453,459,494/Binding site: carboxylate (Asn) (covalent) #status predict
 F:396/Binding site: carboxylate (Asn) (covalent) #status experimental
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 23.3%; Score 570; DB 1; Length 638;
 Best Local Similarity 40.8%; Pred. No. 9e-38;
 Matches 116; Conservative 52; Mismatches 88; Indels 28; Gaps 9;

QY 189 YVRGCGAGHVT-----OCTACGHRGYSRIYGMNLSLQPMQASLOFO---G 228
 Db 362 YEAG--SSGTSRLCKVSSDCTT-----KINRIYGTNSSLGEMPMQVSLQVLSQ 415
 QY 239 YHLGGSVITPLMITTAHCVYDLYLPKSWTIOGVLSL--LDNPAHSLVERKIYHSHY 296
 Db 416 NHMGGSILGRWILFNAHCFDGIYPDVMRIYGIILNLSITNKPFPSSIKELIHQY 475
 QY 297 KPRRLNDIALMKLAGPLTFENMIOPVCLPNSSENPDKVCWTCGATDAGADSPV 356
 Db 476 KMSGSDYDLALIKLPNTLTFEOKPLPSKADNTIYTNVWVGTYRE--RGETQNI 534
 QY 357 LNHAAPVPLISNKC--NHRDVGGIISPSMTCAGYLTGVDSCGDSGCPVCOERLWK 414
 Db 535 LQKATITLVNECQKKRYR---VITKQMICACYKKGSGIDACKDGSGLVCKHSGRWQ 591
 QY 415 LVGATSFICGCAEVNKPQVYTRVTSFLDWMHEON---ERDLKT 454
 Db 592 LVGITSWEGCARKEQPGVYTKVAYEYIDWILEIQSSKERALET 635

RESULT 7 KQMSPL

plasma kallikrein (EC 3.4.21.34) precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
 C:Accession: A36557
 R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Braachpape, L.; Roche
 DNA Cell Biol. 9, 737-748, 1990
 A>Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compar
 A:Reference number: A36557; MUID:91090844
 A:Accession: A36557
 A:Molecule type: mRNA
 A:Residues: 1-638 <SEI>
 A:Cross-references: GB:M58588; NID:q200358; PIDN:AAA63393.1; PID:q200359
 A>Note: part of this sequence, including the amino ends of both the heavy and light c
 C:Comment: This protein, synthesized in the liver, circulates as a noncovalent comple
 are linked by one or more disulfide bonds.
 C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; fibrinolytic; glycoprotein; hydrolase; in
 F:1-19/Domain: signal sequence #status predicted <Sig>
 F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
 F:20-109/Domain: apple repeat <AP1>
 F:110-199/Domain: apple repeat <AP2>
 F:200-289/Domain: apple repeat <AP3>
 F:291-380/Domain: apple repeat <AP4>
 F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>
 F:391-621/Domain: trypsin homology <TRY>
 F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,
 F:127,215,308,396,494/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 22.9%; Score 561; DB 1; Length 638;
 Best Local Similarity 34.8%; Pred. No. 4.7e-37;
 Matches 126; Conservative 61; Mismatches 109; Indels 66; Gaps 12;

QY 132 DMKGHYANVACAOIGPFSYSSDNLRYSSLEGGFREFEVSIDHLLPDDKVTALHSHVYVR 191


```

DB 301 DFGGEELNV-----TFVQAGADVOCETCTKTRICQFF-IYSLLPD-----CKE 342
QY 192 EGC-----ASGHVTL-----QCTACGHRGYSRITVGMS 223
DB 343 EGCKSLRLSTDSGSPRITRYGMOGSSGYSLRLCKLVSDPCTT-----KINARIYGGTNA 397
QY 224 LLSQMPQASLQFQ---GYHLGCGSVITPMTITTAHCYVDLPLPKSWTIQVLSV--L 278
DB 398 SLEGMQVSLQYKIVSQHLCGSGIIGROWVLTAAHCFGCIPIYDVKRIYGLISLSEI 457
QY 279 DNPAPSHLVKIVYHSKYPKRLGNDIALMKLAGPLTFNEMIQVCLPNSSENFPGDKVC 338
DB 458 TKETPSRIKELIHOEYKSEGNVDIALIKLQPLNTYEFQKPICLPSKADNTIYTYNC 517
QY 339 WTSGMGTEDGADSPVNLNHAAPLISNKIC--NHRDVGGIISPSMLCAGLTGCVDS 396
DB 518 WVTGMGTYKE--QGETONILQKATIPVLPNECCKRYRDY---VINKQMLCAGKEGDTA 573
QY 397 CGDSDGSPVLCQERRLMKLVGATSPFGICAEVKNKPGVYTRVTSFLDMIHQME---RDL 452
DB 574 CKDSDGSPVLCVCKHSGKQVLTGITSWEGCGCRKQDPGYTVYKVSFMDVILEKTSQVRL 633
QY 453 KT 454
DB 634 ET 635

```

RESULT 8

KOHUP

plasma kallikrein (EC 3.4.21.34) precursor - human
 N:Alternate names: kininogenin; plasma prekallikrein
 C:Species: Homo sapiens (man)
 C>Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999
 C:Accession: A00921, A37939

R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.
 Biochemistry 25, 2410-2417, 1986
 A:Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four
 A:Reference number: A00921; MUID:86243359

A:Molecule type: protein
 A:Residues: 1-638 <CHU>
 A:Cross-references: GB:M13143; NID:9190262; PIDN:AAA60153.1; PID:9190263
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.
 Biochemistry 30, 2050-2056, 1991

A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of
 A:Reference number: A37939; MUID:91152016

A:Accession: A37939
 A:Molecule type: protein
 A:Residues: 20-27:40-46, 'X', 48, 'H', 50, 'X', 52-70, 'H', 75-76, 'X', 78-80, 103-113, 131-140, 141-
 260-283, 'X', 285, 287-291, 'X', 293-295, 314-317, 'X', 319-320, 321-324, 'X', 329-333, 334-339, 'X',
 525:538-551, 562, 'X', 564-567, 573, 'X', 575-576, 578-583, 'X', 585, 592-604 <MCN>

C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
 C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11
 are linked by one or more disulfide bonds.
 C:Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal r
 ingen and may also play a role in the renin-angiotensin system by converting prorenin i
 C:Genetics:

A:Gene: GDB:KLK3
 A:Cross-references: GDB:127575; OMIM:229000
 A:Map position: 4q35-4q35

C:Superfamily: coagulation factor XI; trypsin homology
 C:Keywords: blood coagulation; duplication; fibrinolytic; glycoprotein; hydrolase; infla
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-638/Product: plasma kallikrein #status predicted <MAT>

F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>
 F:20-109/Domain: apple repeat <API>
 F:110-199/Domain: apple repeat <AP2>
 F:200-289/Domain: apple repeat <AP3>
 F:291-380/Domain: apple repeat <AP4>
 F:391-638/Domain: plasma kallikrein light chain #status predicted <LCH>

F:391-621/Domain: trypsin homology <TRY>
 F:21-104, 47-77, 51-57, 111-194, 137-166, 141-147, 201-284, 227-256, 231-237, 292-375, 322-328, 383

F:127,308,396,453,494/Binding site: carboxylate (Asn) (covalent) #status experimenta
 F:318-347,340-345/Disulfide bonds: #status predicted
 F:390-391/Cleavage site: Arg-Ile (coagulation factor XIIa) #status predicted
 F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 22.1%; Score: 540; DB 1; Length 638;
 Best Local Similarity 35.7%; Pred. No. 2,3e-35;
 Matches 124; Conservative 57; Mismatches 116; Indels 50; Gaps 12;

```

QY 132 DMKGHYANVACAOGLPSPVSSDNLRYVSLGQFREFEVSIDHLLPD-----179
DB 301 DFGGEELNV-----TFVQAGADVOCETCTKTRICQFFIYSLLPD-----CKE 342
QY 180 -----KYTALHSHYVYREGCSGHVTL-----QCTACGHRGYSRITVGMSLSQW 228
DB 352 SMDGSPRI--AYGTQG--SSGYSLRLCKLVSDPCTT-----KINARIYGGTNA 397
QY 229 PMQASLQFQ---GYHLGCGSVITPMTITTAHCYVDLPLPKSWTIQVLSV--L 278
DB 403 PMQVSLQVLTQARHLGCGSLIGHQVLTAAHCFDGLPLQDVKRIYGLISLSEI 457
QY 284 SHLVKIVYHSKYPKRLGNDIALMKLAGPLTFNEMIQVCLPNSSENFPGDKVC 338
DB 463 FSOIKEIITHQNKVSEGNHDIKILQAPLNTYEFQKPICLPSKADNTIYTYNC 517
QY 344 GATEDAGDASPYLNHAAPLISNKICNHRDVGGIISPSMLCAGLTGCVDS 396
DB 523 GFSEKE--KGETONILQKATIPVLPNECCKRYRDY---VINKQMLCAGKEGDTA 573
QY 403 GPLVCOERRLMKLVGATSPFGICAEVKNKPGVYTRVTSFLDMIHQME---RDL 452
DB 580 GPLVCKHNGMWRVLGITSWEGCGARQDPGYTVYKVSFMDVILEKTSQVRL 633

```

RESULT 9

PLPG

plasmin (EC 3.4.21.7) precursor - pig (fragment)
 N:Alternate names: plasminogen

N:Contains: miniplasminogen
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 07-Sep-1990 #sequence_revision 01-Nov-1996 #text_change 18-Jul-1997
 C:Accession: S03733; S03737; A25834
 R:Schaller, J.; Marti, T.; Roesseler, S.J.; Kaemper, U.; Rickli, E.E.
 Fibrinolysis 1, 91-102, 1987

A:Title: Amino acid sequence of the heavy chain of porcine plasmin. Comparison of the
 A:Reference number: S03733

A:Accession: S03733
 A:Molecule type: protein
 A:Residues: 1-560 <SCH>
 R:Brundisholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lerger, W.; Manneberg,
 Eur. J. Biochem. 114, 465-470, 1981

A:Title: Comparison of the primary structure of the N-terminal CNBr fragments of huma
 A:Reference number: S03735; MUID:81212097

A:Accession: S03737
 A:Molecule type: protein
 A:Residues: 1-57 <BRU>
 R:Marti, T.; Schaller, J.; Rickli, E.E.
 Eur. J. Biochem. 149, 279-285, 1985

A:Title: Determination of the complete amino-acid sequence of porcine miniplasminogen
 A:Reference number: A25834; MUID:85203907

A:Accession: A25834
 A:Molecule type: protein
 A:Residues: 450-790 <MAR>

C:Function:
 A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a
 us the walls of the graafian follicle; also activates the urokinase-type plasminogen

A:Pathway: fibrinolysis
 C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homol
 C:Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma; serine pr
 F:1-790/Product: plasminogen #status predicted <PRO>
 F:1-77/Domain: plasminogen-related protein precursor homology (fragment) <PLPGH>
 F:1-77/Domain: activation peptide #status predicted <APM>

F:78-560/Product: plasmin chain A #status predicted <ACH>
 F:84-162/Domains: kringlike homology <KR1>
 F:166-243/Domains: kringlike homology <KR2>
 F:256-333/Domains: kringlike homology <KR3>
 F:358-435/Domains: kringlike homology <KR4>
 F:450-790/Product: miniplasminogen #status experimental <MIN>
 F:461-540/Domains: kringlike homology <KR5>
 F:561-790/Product: plasmin chain B #status experimental <BCH>
 F:561-783/Domains: trypsin homology <TRY>
 F:30-54,34-42,84-162,105-145,133-157,166-243,169-297,187-226,215-238,256-333,277-316,305
 bonds: #status predicted
 F:602,645,740/Active site: His, Asp, Ser #status predicted

Query Match 21.5%; Score 526; DB 1; Length 790;
 Best Local Similarity 41.7%; Pred. No. 3,8e-34;

Matches 110; Conservative 42; Mismatches 86; Indels 26; Gaps 7;

QY 194 CASGHTVLTCTACGHRGYSRIVGNSMLLSQMPQASLOFGY--HLCGGSVITPLW 251
 DB 547 CGRPKVPKPCPA-----RVGCGVSIPHSMPQISLRK-YGHCQGLTISPFW 596
 QY 252 IIRAHGVYDYLKPSVTIOGLVSLDNPAPSHLVKTI--VYHSKTKPRKLGNDIALMK 309
 DB 597 VLTAKHLEKSSSPSSYK-----ILGAHEHYLGEVQETIDVSKLFKEPSEADIALK 650
 QY 310 LAGPLFENMIOPVCLPENSENPFDPKVCWTSGATED--GAGDASPVLNHAAPLISN 367
 DB 651 LSSPAVITDVIAPCLPTPNVADRTACTITGERTKGTGAG----LKEALPLIEN 706
 QY 368 KICNHRDVGIIISPMCAGYLTGVDSGCGSGPLVCOERLMLKLVGATSGICAE 427
 DB 707 KVCRRYELGLGKVSPELCAGLAGIDSCGSGPLVCEPKDKYILQGTSMGLCAL 766
 QY 428 VNKPGVTRYTSFLDMWHEQMERD 451
 DB 767 PNKPGYVRSRVFTWIEETIRRN 790

RESULT 10

JCS759
 brain-specific serine proteinase (EC 3.4.21.-) - mouse

C:Species: Mus musculus (house mouse)
 C:Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 31-Mar-2000
 C:Accession: JCS759

R:Yamamura, Y.; Yamashiro, K.; Tsurinaka, N.; Nakazato, H.; Tsujimura, A.; Yamaguchi, N.
 Biochem. Biophys. Res. Commun. 239, 386-392, 1997

A:Title: Molecular cloning of a novel brain-specific serine protease with a kringlike-like
 A:Reference number: JCS759; MUID:98008848

A:Accession: JCS759

A:Molecule type: mRNA

A:Residues: 1-761 <YAM>

A:Cross-references: DDBJ:D89871

A:Experimental source: brain

C:Superfamily: brain-specific serine proteinase; scavenger receptor cysteine-rich domain

C:Keywords: glycoprotein; hydrolase; serine proteinase

F:85-157/Domains: kringlike-like #status predicted <KR1>

F:163-266/Domains: scavenger receptor cysteine-rich domain homology <SPRC7>

F:166-266,273-372,386-486/Domains: scavenger receptor cysteine-rich #status predicted <SR>

F:513-516/Domains: furin binding #status predicted <FR>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

F:513-516/Domains: trypsin homology <TRY>

QY 104 --EYRCVAVGQON--AVLOVFTASWKTMCSDPMKHYANVACALGFP-----SY 150

DB 381 SPGPPIRLVGNENKBEKREVEFVNGQGTICDDCWTFKHAACICRQGYGPAPRTMAY 440

QY 151 -----VSDNLRVSSLEGOFFREFFSIDHLPDDKVTALHSHV-----YRECCAS 196

DB 441 FGEKGPIMHNNVKTG-----NEKALADCVKODIRHNCRSEDAGVTCIDYLEKASS 494

QY 197 GHVATLTCTACG-----HRRGYSRIVGNSMLLSQMPQASLOFGY--HLCGGSVITPLW 251

DB 495 SGNKEMSSCGGLRLHRR--QKRITIGNSLBRAMQASLRKRSAGDRILCGATLL 552

QY 248 TPIMVITTAHCY-----YDLVLPKSTIOGLVSLDNPAPSHLVKTI 291

DB 553 SSCVWVLTAAHCFKRYGNNSRYAVRGDYHTLVEEPEOEIG-----VOQIV 599

QY 292 YHSKTKPRKLGNDIALMKLACP-----LTFNMIDPVLPSNENFPD--GKVCITSGWA 345

DB 600 IHRVYRDPDSYDIALVRLQGPGEQCARLSTHYLPACLPIMRER--PKTASNCBITGWD 658

QY 346 TEDGAGDA--SPVLNHAAPLISNKCINHRDVGIIISPMCAGYLTGVDSGCGDSG 402

DB 659 T-----GAYSTLTQQAAPLIPKPFCKR--YKLPFGKMLCAGNLDENRVRVSCGDSG 712

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

QY 403 GPLVCOE--RLMLKLVGATSGICAEVKNKPGVTRYTSFLDMW 444

DB 713 GPLMCEKPDSEWVYVYVTSMGCGVADTGVYRVPAFVPMI 755

Db 264 TAAHCLERSRSSTKYKVLGTHHELRLAAGAQDD -VSKLEFESRA -DIALKLKSSP 319
OY 314 LTFNMIQPVCLPNSEENFPDGKVCVWTSGMGATEDGADSPVNLNAAVPLLSKTKCHR 373
Db 320 AITQNVTPACLPPADYVVAWMAECFVTGMEGTODSSN -AGVLEADLPYIENKVCRRY 377
OY 374 DVGGIISPSMCAAYLTGVGDCGCGDSGGLVCCERLRMLKVGATSTGIGCAEYVNPFGV 433
Db 378 EYLNQRVASTELCAGHLVGVDCGDSGSGPLVCFEKKDYILQGYTSMGLGCAEPNPKGV 437
OY 434 YTRVTSFLDWIHEOME 449
Db 438 YVRVSSFIWIERIMO 453
RESULT 12
PLMS
plasmin (EC 3.4.21.7) precursor - mouse
N:Contains: angiotatin; plasminogen
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1991 #sequence revision 01-Nov-1996 #text_change 18-Jun-1999
C:Accession: A38514; S48202; S48203
R:Degen, S.J.F., Bell, S.M., Schaefer, L.A.; Elliott, R.W.
Genomes 8, 49-61, 1990
A:Title: Characterization of the cDNA coding for mouse plasminogen and localization of t
A:Reference number: A38514; MUID:91184812
A:Accession: A38514
A:Molecule type: mRNA
A:Residues: 1-812 <DEG>
A:Cross-references: GB:J04766; NID:g200402; PIDN:AAA50168.1; PID:g200403
R:Lijnen, H.R.; Van Hoel, B.; Beelen, V.; Collen, D.
Eur. J. Biochem. 224, 863-871, 1994
A:Title: Characterization of the murine plasma fibrinolytic system.
A:Reference number: S48202; MUID:95010076
A:Accession: S48202
A:Molecule type: protein
A:Residues: 20-25 <LID>
A:Accession: S48203
A:Molecule type: protein
A:Residues: 22-27 <LID>
C:Comment: Plasminogen is synthesized by the kidney and is present in plasma and many o
C:Comment: Plasminogen is converted into plasmin by plasminogen activators, both plasmin
immediately after dissociation from the clot. In the presence of the inhibitor, the activ
e inhibitor, the activation involves also removal of the activation peptide.
C:Comment: Stromelysin I (see PIR:KCMSI) acts on plasminogen to produce angiotatin. Ic
eeful in treating solid tumors.
C:Function:
A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va
as the walls of the graafian follicle; also activates the urokinase-type plasminogen act
A:Pathway: fibrinolysis
C:Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology
C:Keywords: angiogenesis inhibitor; blood; duplication; fibrinolysis; glycoprotein; hyd
F:1-96/Domain: plasminogen-related protein precursor homology <ELPH>
F:20-812/Product: signal sequence #status predicted <SIG>
F:20-812/Product: plasminogen #status predicted <PRO>
F:20-96/Domain: activation peptide #status predicted <APT>
F:79-466/Product: angiotatin #status predicted <AST>
F:97-581,582-812/Product: plasmin #status predicted <AMT>
F:97-581/Domain: chain A #status predicted <ACH>
F:103-181/Domain: kringle homology <KR1>
F:185-262/Domain: kringle homology <KR2>
F:275-352/Domain: kringle homology <KR3>
F:377-454/Domain: kringle homology <KR4>
F:481-560/Domain: kringle homology <KR5>
F:582-812/Domain: chain B #status predicted <BCH>
F:582-805/Domain: trypsin homology <TRY>
F:49-73,53-61,103-181,124-164,152-176,188-262,188-316,206-245,234-257,275-352,296-335,32
bonds: #status predicted
F:78-79/Cleavage site: Glu-Asn (stromelysin I) #status predicted
F:136,308/binding site: carbohydrate (Asn) (covalent) #status predicted
F:466-467/Cleavage site: Thr-Val (stromelysin I) #status predicted
F:581-582/Cleavage site: Arg-Val (plasminogen activator) #status experimental

```

F:624,667,762/Active site: His, Asp, Ser #status predicted

Query Match          20.7%   Score 507.5;   DB 1;   Length 812;
Best Local Similarity 39.8%   Pred. No. 1,26-32;
Matches 106; Conservative 40; Mismatches 91; Indels 29; Gaps 7;

QY 194 CASGHVYVLTQCTACGHRGYSRSRTIGVGNMISLSQMPQASL--QFGCHLCGGSVIPLM 251
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 568 CGKPOVEPKK-----PGRVVGCCVAPNHPSPWISLRTPTGQHFCGGLTIAEW 618
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 252 IITAAHCYVDYDLRPSWITQVG---LVSLLDNAPSHLVKRIYHSHKPKRKLGNIAL 307
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 619 VLTAAHCLKSSRPREFYVILGAAHEEYIRGLDVGELS--VAKLLLEPR-----NRDIAL 670
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 308 MKLAGEPLTFNEMIQPVCLPNSDENPDDCKVYCWTSQWATED--GAGDASPLNNAAPLI 365
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 671 LKLSPPATITKIVLPACLPSPNYVADRTICYITIGWGTGTFAGR---LKEADQLPVI 726
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 366 SNKICNHNDVYGGIISPMLCAGLYLTGVDSCQSGDSGPLVCOERRLMKLVGATSFQIC 425
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 727 ENKVCNREYELNNRKKSHLCAQLAGVDSQGDSSGPLVCFENDKITYILOGVYSWGILGC 786
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 426 AEVNEKPVYTRVTSFLDIHEOMERD 451
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 787 ARPNEKPVYTRVTSFLDIHEOMERD 812

RESULT 13
acrosin (EC 3.4.21.10) precursor - rat
N:Contains: proacrosin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 22-Jun-1999
C:Accession: S18407; S30037; A56620
R:Kleem, U.; Flake, A.; Engel, W.
B:Title: Rat sperm acrosin: cDNA sequence, derived primary structure and phylogenetic
A:Accession: S18407
A:Molecule type: mRNA
A:Residues: 1-437 <KLE>
A:Cross-references: EMBL:X59254
R:Kleem, U.; Flake, A.; Engel, W.
submitted to the EMBL Data Library, April 1991
A:Reference number: S30037
A:Accession: S30037
A:Molecule type: mRNA
A:Residues: 1-254, 'ICDR', 259, 'DHEL', 264, 'GRIC', 269-437 <KLE>
A:Cross-references: EMBL:X59254; NID:957282; PIDN:CAAL1947.1; PID:957283
R:Kremling, H.; Flake, A.; Adam, I.M.; Radtke, J.; Engel, W.
DNA Seq. 2, 57-60, 1991
A:Title: Exon-intron structure and nucleotide sequence of the rat proacrosin gene.
A:Reference number: A56620; MUID:92199245
A:Accession: A56620
A:Status: preliminary
A:Molecule type: DNA, mRNA
A:Residues: 1-254, 'ICDR', 259, 'DHEL', 264, 'GRIC', 269-437 <KRE>
A:Note: sequence modified after extraction from NCBI backbone
A:Note: sequence extracted from NCBI backbone (NCBIN:89436, NCBIN:89439, NCBIN:89447,
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm; zymogen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-43/Domain: acrosin light chain #status predicted <LCH>
F:44-286/Domain: trypsin homology <TRY>
F:49-437/Domain: acrosin heavy chain #status predicted <HCN>
F:22-211/Binding site: carboxylate (Asn) (covalent) #status predicted
F:25-155,29-163/Disulfide bonds: #status predicted
F:74-90/Disulfide bonds: #status predicted
F:89,143,241/Active site: His, Asp, Ser #status predicted
F:178-247/Disulfide bonds: #status predicted
F:210-226/Disulfide bonds: #status predicted
F:237-267/Disulfide bonds: #status predicted

```


